DOCUMENT RESUME

ED 340 373 IR 053 865

AUTHOR	Chiang, Katherine; And Others
TITLE	INFeRS: Interactive Numeric Files Retrieval System. Final Report.
INSTITUTION	Cornell Univ., Ithaca, N.Y. Univ. Libraries
SPONS AGENCY	Department of Education, Washington, DC.
REPORT NO	R197-A80324
PUB DATE	91
NOTE	155p.
PUB TYPE	Reports - Descriptive (141)
EDRS PRICE	MF01/PC07 Plus Postage.
DESCRIPTORS	Academic Libraries; Access to Information; Agriculture; *Computer System Design; Database Design; Higher Education; Information Retrieval; Information Systems; Man Machine Systems; *Numeric Databases; Online Systems; Specifications; *Systems Development
IDENTIFIERS	Cornell University NY

ABSTRACT

In 1988 Mann Library at Cornell University proposed to develop a computer system that would support interactive access to significant electronic files in agriculture and the life sciences. This system was titled the Interactive Numeric Files Retrieval System (INFeRS). This report describes how project goals were met and it presents the project's conclusions, including recommendations on how numeric file interfaces should be structured. The first of six sections presents the project chronology and indicates the months during which significant project events occurred, including personnel recruitment, technical hardware and software installations, phase completions and testing. Background information provided in the second section includes information on the personnel who participated in the project; the selection, acquisition, and installation of the hardware and software used; and the institutional environment, including descriptions of the library and the computing and networking resources of the user group. The third section covers the selection, preparation, and documentation of the datasets used in the project. Database design issues are discussed in the fourth section, including the search and file transfer modules and such interface design issues as the screens, search sequence, and indexing modules. The fifth section provides a detailed description of the system, and future plans for the interface are presented in the sixth, together with a summary of the major issues and recommendations and conclusions. Documents related to the project are appended, including two requests for proposals (one for a minicomputer and the other for database management software), sample data output, and full codebook. (MAB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED340373

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
● Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

INFeRS: Interactive Numeric Files Retrieval System

Final Report
R197 A80324
to the

U.S. Department of Education
Research and Demonstration Project for
the Sharing of Library Resources

Katherine Chiang
Computer Files Librarian
(607) 255-2199
Mann Library
Cornell University
Ithaca, NY 14853-4301

IR053865

BEST COPY AVAILABLE

Table of Contents

Introduction *Katherine Chiang* 1-3
  Goals of the Project 1
  Final Report 2

Chronology *Katherine Chiang* 4-5

Background *Katherine Chiang, Leslie McLane* 6-13
  Insitutional Context 6
  Personnel 8
  Technical 10

Data 14-19
  Data Selection *Katherine Chiang* 14
  Data Preparation *Katherine Chiang, Tom Randolph* 15

System Creation 20-39
  Introduction *Leslie McLane* 20
  Project Phases *Leslie McLane* 20
  System Specifications, Design and Development *Leslie McLane* 21
  Query Construction and Execution *Leslie McLane* 24
  Database Design and Construction *Leslie McLane* 26
  Data Problem Resolution *Leslie McLane* 27
  Subject Indexing *Marijo Wilson* 29
  File Transfer *William Garrison* 32
  Interface Design *William Garrison* 34

System Description *Katherine Chiang* 40-66
  Transactions 40
  Detailed Description 40
  Helps 63

Conclusion *Katherine Chiang* 67-72
  Still to do 67
  Summary of Major Issues 69
  Research and Demonstration to Reality 71

Appendices

Introduction

The goal of INFeRS is to simplify the confusion
—Tom Randolph, project statistician

Goals of the Project

In 1985 B. C. Carter, staff director of the Numerical Data Advisor Board of the National Research Council/National Academy of Science, wrote of the need for "the development of sophisticated, yet user-friendly software not just to address a single specialized database, but increasingly also to be used in gateway and networked data systems that provide numerical structure search, calculational, and statistical capabilities." (Carter, 1985)[1]

In 1988, responding to that need Mann Library at Cornell University proposed "to develop a computer system supporting interactive access to significant electronic files in agriculture and the life sciences in four phases: (1) system design and implementation, (2) system testing and enhancement, (3) project evaluation on the cornell campus, (4) provision of access to the system from remote locations." (Mann Library, 1988) [2]

The creation of such an Interactive Numeric Files Retrieval System (INFeRS) will, we hope, simplify the process of collecting a data subset from a large file. Learning how a dataset is organized, and learning software that is capable of extracting a useful subset is difficult and can be very confusing. Thus our motto for INFeRS was — To Simplify the Confusion.

The goals of the project, as stated in our proposal of May 1988, were based on a three-year timeline, at full funding. As a result of our negotiations with the Department of Education we were funded for a two-year project, with our goals becoming what we had planned to do in the first two years of the three-year project we originally submitted.

Those goals were:

—to assemble a computer system capable of retrieving a minimum of 250 megabytes of data, with access to that system from selected microcomputers in the Library.

—to select four datasets to load onto the system, selection includes identifying faculty sponsors and a focus group of interested researchers to identify the functions the system would need.

—to design and create an interactive data retrieval system.

—to create documents that describe the process and the final system.

All of the goals were met.

We have assembled a computer system with approximately 1.5 gigabytes of storage. That process is described in the Background Section.

We selected four datasets and assembled the related faculty sponsors and interest groups and interviewed them for advice on system functions. That process is described in the Data Section.

---

[1] Carter, B.C. "Numerical Databases: Their Vital Role in Information Science. Part II: A Call to Action." *ASIS Bulletin*, April/May 1985, p. 10

[2] Mann Library. "Information Resources on the National Network" U.S. Department of Education, Title IID Grant Application. May 1988. p. 1

We designed and created two interactive data retrieval systems: a simple one to give us experience in the capabilities of the machine and the software, and a full-fledged system. That process is described in the System Section.

This report fulfills our final goal of documenting the process and the final system.

Goals of the Final Report

In this document we report on the activities funded by our Title IID grant. The document includes the details on how our project goals were met. More importantly it records our conclusions from and generalizations on the project, including recommendations on how numeric file interfaces should be structured. People, or institutions, interested in creating broad-audience numeric file interfaces face a challenging task. This report focuses on that information about our experiences and conclusions that can assist others. It should be of interest to any library planning to create a similar interface, or some other sort of access to data, and those using numeric interfaces.

Structure

The report is divided into six sections: Project Chronology, Background, Data, System, Conclusions, and Appendices.

PROJECT CHRONOLOGY The chronology simply lists the dates on which significant project events occurred, including personnel, technical hardware and software installations, phase completions and testing.

BACKGROUND This section includes discussions of the personnel who participated in the project, the selection, acquisition and installation of the hardware and software used, and the institutional environment, including descriptions of the library, the computing and the networking resources of the user group.

DATA This section covers the selection preparation, and documentation of the datasets used in the project.

SYSTEM This section discusses the database design issues, including the search and file transfer modules, and the interface design issues, including the screens, search sequence, and indexing modules.

CONCLUSIONS The future plans for the interface are described, as well as a summary of our recommendations and conclusions after working with the creation of a numeric files interface.

APPENDICES Documents relating the project are assembled in the Appendices. See the Table of Contents for a complete list.

Project Chronology

1988

May	Applied for grant
September	Notified grant received

1989

January	Project began
March	Hardware RFP issued
May	Programmer job search
June	Faculty interviews
	Hardware Vendor demonstrations
September	Programmer hired
	Hardware ordered
October	Statistician job search
November	Hardware installed
December	Software RFP issued

1990

January	Statistician hired
	Software vendor presentations
February	Interface designer job search
	Software ordered
March	Crop Estimates data checking
	Software installed
	Programmer training in Informix and C
April	Phase I begins
June	Phase I complete
	Background work on National Resources Inventory, query logic, and screen development
July	Interface designer hired
September	Phase II begins
October	Informix consultant visit
November	User reviews of screens and indexes

1991

January	Testing of Phase II interfaces
	Phase III begins

February	Phase II complete
April	Phase III complete
	Document programming
	Debugging program
May	Programmer finishes
June	Instruction packet for users created
	Statistician finishes
July	Title IID project ends, INFeRS moves into maintenance mode
October	Final Report submitted to the Department of Education

Background

Institutional Context

Organization

CORNELL Cornell is a private and state funded coeducational university, founded in 1868. The University has approximately 1,600 faculty, and 6,100 graduate students, and 12,700 undergraduates on the Ithaca campus, enrolled in over 100 degree programs.

Cornell's physical location is perhaps a cause for Mann Library's interest in projects to provide electronic access to physically distant resources. The University is best described by the locally available T-shirts emblazoned with *Ithaca, New York, Centrally Isolated.* Ithaca is in upstate New York, it is a four-hour drive to New York, seven to Boston and Washington. The implications for the students and faculty of Cornell are obvious; all the resources necessary for research and instruction must be locally available or accessible via computers. Driving to a nearby library is not an option.

CORNELL UNIVERSITY LIBRARY The University is served by a network of eighteen libraries loosely defined by their subject areas. The libraries range in size from the Africana Library, with 14,000 volumes, to Olin library, with almost 3 million volumes.

Mann Library has the second largest collection on campus, with over 600,000 volumes. It serves the College of Agriculture and Life Sciences, the College of Human Ecology, the Division of Biological Sciences, and the Division of Nutritional Sciences. The collection contains materials on agriculture, life sciences, and social and behavioral sciences. Our primary users are the 536 faculty, 1,438 graduate students, and 4,354 undergraduates of those Colleges and Divisions.

Computing Environment

CAMPUS Campus computing is supported by College and Departmental technical staff and Cornell Information Technologies (CIT), the campus-level systems support. CIT administers the mainframes and some of the many minicomputers on campus. They have departments for systems support, programming, and user services.

Most of the administrative computing is handled by two IBM mainframes. Student do their computing on a CIT administered VAX and on machines in the microcomputer centers scattered throughout the campus in classroom buildings and dormitories. A significant percentage of students own their own machines, Macintosh being the system of choice.

The University also administers the Cornell National Supercomputing Facility.

LIBRARY COMPUTING The Cornell University Library uses the NOTIS software for their automated systems.

Mann Library offers mediated searching of bibliographic databases. In addition, librarians teach students how to do their own searching and the library offers an end-user search service and has loaded bibliographic databases onto a local multi-user machine for distributed searching. Almost all the libraries on campus have compact disk bibliographic databases, and several also have text or numeric databases on CDs.

Our student patrons range from freshmen who have no idea what a bibliographic citation is, and have never used a computer, to freshmen who have been searching DIALOG since high school. The faculty show the same range of skills and expertise in using computers.

Target Audience

We chose the datasets based on a set of criteria. Those choices then determined the potential audience for the data. Details on the criteria for the datasets are given in the Data Selection section. Stated broadly, our audience was anyone who might want the data we put on the system. That rather bold definition was refined in the course of our discussions about how the system should function, and what we learned from our patrons.

We categorized prospective patrons into several groups based on several characteristics.

Type and amount of data needed:
—patrons who need a few summary statistics,
—patrons who need a large set of summary statistics,
—patrons who need a small set of data, and
—patrons who need large datasets.

Level of knowledge of the data available:
—patron who does not know where the data might come from,
—patron who knows which dataset contains their data, but has not used the dataset, and
—patron who is an expert on the dataset.

Level of knowledge of computers:
—patron who rarely uses computers for anything,
—patron who uses computers for word processing and other basic tasks,
—patron who uses other, perhaps specialized, applications software in the course of research, and
—patron who programs.

From our interviews with patrons we found prospective users of the data with almost every permutation of the above characteristics: expert computer users who were also experienced users of the datasets we had chosen, computer novices who had used the data (albeit in print form) for years, middle computer users who need a couple of summary statistics, etc.

This profiling was a prerequisite to our system design. The System Section includes details and examples of how what our patrons could do (and would want to do) influenced decisions about the interface.

Personnel

The system was created by five members of Mann library, with the support of other members of the Library and the Cornell community.

Organization, administration

The five members of the project team came from three of the four divisions in the library. Those four divisions are:

—Administrative (Office of the Director, personnel, budget, and the Information Technology Section)
—Technical Services (acquisitions and cataloging)
—Collection Development (selection, preservation)
—Public Services (reference, circulation, interlibrary loans, reserves.)

Project team

Katherine Chiang, computer files librarian, Public Services, assigned 35% to the project. Katherine was the primary researcher, responsible for the progress and administration of the project. She organized the structure of the project, the meetings, and made assignments. She was responsible for reporting to the Department of Education and liaison with the Head of Public Services. She was also responsible for providing users access to the data during the creation of the system.

Leslie McLane, programmer, Information Technology Section, funded from the grant and assigned 100% to the project. Leslie was the programmer and database designer. She took the preliminary specifications for the system, developed the detailed function specifications and determined how they would be implemented in Informix. She programmed the system, all phases. (See the Systems Section for details of her work.)

Bill Garrison, interface designer, Public Services, assigned 40% to the project. He started work in the library July 1990, in time to work on the design on the screens and interface for the second phase of the project. He was responsible for the file transfer process and the technical issues surrounding the communications programs that could be used to access the system. (See the Systems Section: Interface Design for details of his work.)

Marijo Wilson, cataloger, Technical Services, assigned 10% to the project. Marijo was responsible for how the documentation and variable/field value information would be displayed to the user. She researched missing information, and designed and created the Subject and Field

Indices. (See the Systems Section: Subject Indexing for details of her work.)

Tom Randolph, Statistician, Public Services, funded by the grant, part time, 15-20 hours per week on the project. Tom was responsible for the data profile/preparation. He checked the data for errors and determined the details of their structure, e.g., how the data were coded, including their idiosyncracies. He also worked on access issues and interface design, contributing the viewpoint of an experienced data user. (See the Data Section: Preparation for details of his work.)

Team structure

The project team met weekly to report on progress and discuss issues needing resolution. In the early stages of the project, when many basic decisions on staff involvement and project scope were under discussion, the Heads of the Public Services and the Information Technology Section also attended some of the meetings. Team members also met in twos and threes on an ad hoc schedule to work on specific modules.

The team structure worked well. The project leader made project assignments based on the time assignments for each team member. Each team member reported on their project activities to their corresponding Division head. The project leader met regularly with the Director of the library.

The programmer was assigned to the Information Technology Section, even though the majority of the project team was in Public Services, because previous programmers working in Public Services found the absence of daily interaction with other programmers a handicap.

Other Support

The technical and networking issues relating to Albert were handled by Tim Lynch, Network Manager, ITS, and John Hood, Systems Programmer, ITS. John Hood also helped Leslie install the Informix program. With Leslie's departure at the end of the project, the responsibility for the maintenance of the system was transferred to Bill Fenwick, Programmer, ITS.

Howard Curtis, head, ITS, and Susan Barnes, head, Public Services, participated in the user interviews described in the Future Plans section.

Jo Jaynes, Acquisitions Coordinator, Technical Services, was responsible for the acquisition of the datasets loaded into the system. Mary Kelly, Accounts Supervisor, Administration kept the local financial records on the funds provided by the Department of Education.

Test groups

Several different user tests were run during the development of the system. Test subjects were recruited from several different groups.

- —Faculty and graduate students from the College of Agriculture and Life Sciences, identified in the initial interviews.
- —Public Services professionals of Mann Library.
- —Members of the Information Technology Section of Mann Library.
- —Members of the Interactive Media Group of the College of Agriculture and Life Sciences.

Vendor support

We received support from both Hewlett-Packard and Informix. One of the items negotiated into our contract with Informix was two days of consultant help, to be scheduled when we needed it. Paul Wolmering came to Ithaca for two days in October 1990 and worked closely with Leslie answering her questions and reviewing the database design for the second phase of the project.

Technical

Acquiring the computer and software

The acquisition of the minicomputer and database software was handled by the Information Technology Section with Howard Curtis, section head, in charge. Two similar but separate RFP bid processes were run.

MINICOMPUTER The minicomputer acquisition process began in January, with the start of the grant. Howard Curtis wrote the RFP, which was issued in March of 1989. (Appendix A) He assembled a review committee from various departments on campus.

Howard Curtis, chair, Head, ITS, Mann Library
Bill Fenwick, Programmer/Systems Analyst, ITS, Mann Library
Susan Barnes, Head, Public Services, Mann Library
Kathy Chiang, Computer Files Librarian, Public Services, Mann Library
Tom Boggess, Assistant Director, Systems: Computer Systems Services (systems programming for multi-user systems), CIT (previously manager of computer resources, Cornell Institute for Social and Economic Research)
Larry Fresinski, Associate Director, Workstation Resources, CIT
Carol Lambert, Assistant Director, Resource Services, CIT

The group met several times: to review the RFP and how the evaluation should proceed; to review the written proposals to determine which vendors would be asked to make presentations; and to review the presentations and make a recommendation for acquisition.

Three vendors made presentations in June 1989: DEC, Sun, and Hewlett Packard. Based on the information presented in the written proposals, the presentations, and further research, a Unix based, Hewlett Packard HP9000 825S system was selected. After negotiations with Hewlett Packard, the system was ordered in the first week of September, received and installed in November 1989 and eponymously named Albert, after the dean who gave his name to the library.

As of September 1991 Albert consists of a Hewlett Packard 9000 825S with 24 megabytes of memory and 1.6 gigabytes of disk storage, configured for 32 simultaneous users. The system includes a 9-track magnetic tape drive. It runs HP-UX, Hewlett Packard's version of Unix.

DATABASE MANAGEMENT SOFTWARE **[Leslie McLane]** The database selection process began with the decision to use relational database technology as the data management tool primarily because the characteristics of SQL (Structured Query Language) lend themselves well to the ad-hoc nature of INFeRS. Relational technology is, by design, known for its ability to perform well on "on-the-fly" data requests. It is designed to execute unstructured, changing queries in the most efficient (quickest) way possible. This is accomplished by moving the burden of deciding "how" to find the data requested from the user to the software. Due to the power of relational technology and the idea that INFeRS was being developed to be a vehicle to access and subset data, not to actually perform any statistical analysis, it was chosen over popular statistical software solutions such as SAS.

The list of potential database vendors, therefore, was narrowed to RDBMS's (Relational Database Management Systems) that were supported on the HP 9000. The initial list consisted of eight possible vendors and was reduced to a pool of four (Informix, Ingres, Oracle, and Sir) based on evaluations of responses to a detailed Request for Proposal. (Appendix B) The RFP outlined functional criteria vital to the project, and requested responses detailing if and how each need was supported. The four vendors were invited to give an on-site demonstration and presentation of their database products, emphasizing features that met our needs. The members of the evaluation committee were:

Leslie McLane, Programmer, ITS, Mann Library
Howard Curtis, chair, Head, ITS, Mann Library
Kathy Chiang, Computer Files Librarian, Public Services, Mann Library
Marijo Wilson, Cataloger, Technical Services, Mann Library
Tom Randolph, Statistician, Public Services, Mann Library
Tom Dimock, Assistant Director, Technology: Information Resources (database and file server technologies), Cornell Information Technologies
Wilson Manik, Director of Networking, Office of Planning and Information Studies, College of Human Ecology
Andrea Beesing, Applications Programmer, Division of Nutritional Sciences
Sharon Bushart, Applications Programmer, Division of Nutritional Sciences

Major considerations

4GL In addition to database management capabilities, we were seeking a software product that also provided a Fourth Generation Language (4GL) having both built-in 4GL tools as well as a robust programming language. 4GL's are becoming very popular because they provide application development tools that allow quick implementation of standard application components (menus, browsing environments, data entry screens, etc.) without the need for extensive coding. With a 4GL, it is possible to develop quick prototypes of different application scenarios and ideas. This was obviously a very attractive feature because we were exploring new areas and wanted to be able to concentrate our efforts on testing various models without getting bogged down in lots of programming. A Third Generation Language (like C) would require extensive coding to achieve the same results.

C Although all of the RDBMS's evaluated had some sort of programming language and 4GL "toolbox", we were relatively confident that not all of our project needs would be satisfied by any given product. Therefore, we wanted to be sure that the product chosen would be able to communicate with external routines written in a third generation programming language, particularly C.

SQL For INFeRS to perform the way it was initially envisioned, the system had to allow a user to base a search on any variable in a file in any way needed, and select an output format appropriate for loading search results into other data analysis software packages. Therefore, any variable was a potential target for generating filter conditions and/or output file specifications. To perform this task, the RDBMS had to be capable of dealing with a search statement that would change dramatically every time the program was used. Specifically, it had to be capable of creating and executing "dynamically generated SQL statements" — SQL created entirely from scratch.

Although the majority of the evaluated RDBMS's could handle this requirement in some way, a large question remained concerning whether a dynamically-generated SQL statement could be created, executed and controlled within a 4GL program, or if it had to be passed to and executed by a 3GL program (like C). For purposes of displaying results to the screen as well as controlling output to a file, we felt it was necessary to handle the execution of a query from within the 4GL environment. In addition, it was necessary that no limitations be placed on the complexity or length of any given search by constraints of the 4GL. A few of the RDBMS's allowed "dynamic" SQL to be generated by a 4GL program but placed constraints on the total number of variables and constraints on each variable that could be part of the filter condition.

DATA Although the composition of the files identified for loading into INFeRS were primarily coded and numeric, it was necessary for the chosen database environment to be capable of handling searches based on coded, numeric and textual data. Because target search fields cannot be predicted (because they are defined by the user during a session), we were looking for a database manager that was efficient in dealing with ranges of values for numeric fields (production > 1000), specific alpha-numeric entries for coded fields (commodity = '101999' or '101199') as well as substring searches on longer text fields (the word 'irrigation' within a variable name).

Because the data were likely to have occurrences of "missing" or "unknown" data values, the product had to be able to distinguish null values from spaces and zeros.

VT 100 The INFeRS program was being developed to run on the HP and to be accessed by a host of end-user workstations via the campus network or telecommunications devices. To accommodate all users, it was necessary to provide "lowest common denominator" access to the system. Therefore the software had to support VT100 terminal emulation.

Acquisition

The RFP was distributed in early December 1989 with a thirty day response date for vendors. Written bids were evaluated in January. The presentations took place in late January and early

Feburary of 1989. We selected Informix in February and the program was ordered and installed in March.

Installation

The minicomputer was installed in a basement room of the library, sharing the space with old agricultural reprints. New electrical circuits were installed, the room was air conditioned and an Ethernet connection was added. The Ethernet was an extension from the cabling installed in 1987 when an asynchronous terminal network called Sytek was installed in order to access the Cornell Online Catalog.

> The Ethernet, which is a logical bus topology based on a physical star, has its central point of concentration in the telecommunications closet on the first floor of the library. . . . .One strand of cable leads down to the computer in Room 3/4. Eleven other strands of cable radiate out to staff work areas on the second and third floors of the library and to its ITS office area. . . . . These radial strands are concentrated in a "DEMPR" [multi-port repeater] in the telecommunications closet. The entirety of this library Ethernet network gains access to the campus backbone network via a Cornell AT-gateway ("Cogger Box") also located in the telecommunications closet. Also attached to the AT-gateway, and consequently to the campus backbone, are the Ethernet-based Novell network in the Mann Microcomputer Center, and the Appletalk network of Macintoshes in our Public Services Department. The connection to the campus backbone TCP/IP network (CIT Net) is achieved with a 1-Megabit per second Omninet link.
>
> This new networking arrangement provides Mann Library's staff workstations and central information servers with direct access to the campus backbone and to the evolving Internet. (Curtis, 1990)[1]

In addition to the Ethernet access, a telephone line was installed to allow modem access, and four Sytek lines were added to allow users with access to modems, or a Sytek connection, to log onto the minicomputer.

Thus Albert is accessible to anyone with a computer with access to the Internet, Sytek, or a telephone modem. All of the staff machines in the Library can access Albert, as well as all of the public access IBM and IBM-compatible microcomputers in the library.

---

[1] Curtis, Howard Information Technology Section Annual Report 1989-90. Mann Library unpublished document.

Data

Data Selection

Criteria

Data sets were selected to meet several goals. We wanted to guarantee they would be used by our patrons, so we chose datasets in agriculture and the social sciences. For each possible dataset we contacted faculty who would be likely to use the data to see if they would be interested in using them, and participating in the project. Participation involved allowing us interview them on their data use and their using the system in controlled tests.

Since our intent is to make these resources available over national networks we wanted datasets of national interest. We deliberately avoided local databases for this reason. Furthermore, since we were creating a system to subset data, we chose datasets that were either so large or diverse that they would rarely be used in total.

Finally, we chose datasets that were public access, or where network access could be negotiated.

Datasets

Using the criteria above we identified five datasets: The commodity trading statistics (either from the Chicago Board of Trade, or a commercial data source), U.S. Crop Estimates — county (from the National Agricultural Statistics Service of the U.S. Department of Agriculture), the National Resources Inventory (from the Soil Conservation Service of the U.S. Department of Agriculture), the Health and Nutrition Examination Survey (from the Center for Health Statistics, National Institutes of Health), and the Toxic Release Inventory (from the Environmental Protection Agency.)

We acquired and profiled each dataset. With that profile and the information from our interviews with interested researchers, we decided to load the datasets in the following order:

CROP ESTIMATES — COUNTY FILE We decided to use this file for our first phase because we thought it was a relatively small, simple file. We were to learn that simple is not an adjective to be used with datafiles. Details of that learning process are included in the next section on Data Preparation.

NATIONAL RESOURCES INVENTORY For the second phase of the system design we wanted a dataset that would present a wider range of challenges. The NRI is a larger file with more fields in the record and, as we were to discover, many types of fields. Solving the access issues for each variant field in the NRI gave us precedents to be used when loading subsequent files.

COMMODITIES TRADING This dataset is very similar in simplicity to the Crop reporting board. We felt it would be a good exercise to see how quickly a small, clean dataset could be loaded once the system was created.

TOXIC RELEASE INVENTORY (TRI) The TRI is assembled from the forms sent in by the companies so that many of the fields have non-standardized entries. Each record is a mix of free text, numeric, and coded fields. We felt the indexing for this dataset would be quite involved, and we wanted the experience of working with the less complicated datasets before we loaded the TRI.

HEALTH AND NUTRITION EXAMINATION SURVEY (HANES) For this file we received valuable information from the researchers who work with this file regularly. They alerted us to the complexities of working with this dataset. Each record/interview represents a certain portion of the population, but those sample weights vary with each observation. An interface that failed to take that weighting into account would not be particularly useful. For that reason we put this dataset last on the list, thinking we would need all the experience we could get before tackling it.

Data preparation

Note: The content and organization of the following section is the work of the statistical consultant. It has been translated into narrative form from his notes.

Access to the data

The datasets used for the project were acquired through the normal library procedures. The Acquisitions Department ordered the file and checked it in. The Cataloging Department cataloged the file into our online catalog and then gave the file to the project team.

In each case we received the files on magnetic tape. Those tapes were submitted to the CIT tape library, assigned numbers and shelved in the collection. We made a working copy of each file; that copy was used for the data preparation. We ran the data checks on the CIT machine because SAS is loaded on that machine. Albert (the Mann Library computer) does not have a SAS licence. After the data was 'cleaned' it was transferred over to Albert.

Getting to know the data

The first file to be loaded was the Crop Estimates — county from the U.S. Department of Agriculture. We decided to examine this dataset in as detailed a form as we could for three reasons.

First, carrying data preparation to its highest level would allow us to develop a strategy for data cleaning to be used on the other files. Only by 'overkill' would we be able to identify the levels of preparation possible. Only by identifying those levels would we be able to decide on the appropriate level for a 'routine' loading of a dataset.

Second, we also needed to identify the characteristics and problems in the dataset that the programmer needed to know to load the file, and to facilitate the proper design of the database and the interface. Finally, we wanted to identify characteristics and problems about the dataset that the user would need to know.

Therefore, for the Crop Estimates file we followed quite an involved process including an initial verification of the dataset, identification of the physical parameters of the data and any special characteristics. We looked for unreasonable and unreadable data, problems with the uniqueness of record identifiers, and attempted to gain a sense of how the data were generated.

Initial verification

The first thing we determined was whether we had the right data. Did they correspond to the description in the user documentation we received. This was done through a simple printout of the first few records in the file, and SAS descriptive statistics on number of records, etc. By running additional checks we were also able to define the physical parameters of the file.

Special characteristics

Special characteristics of the data need to be identified and explained. How are missing values and blank fields to be interpreted. Were the values not collected?, not available?, not applicable? Were they incorporated into some other aggregate observation? Are they actually a zero value? Close reading of the user manual, and a comparison to printed sources are usually enough to answer these questions.

Another issue is the interpretation of zeroes. Does a zero in a field actually mean a zero value, or an insignificant value. Are they used as blank fillers, or when the value is not available or applicable? Again consultation of the user manual and the print sources is usually enough to resolve these questions.

In the Crop Estimates file we encountered categories of observations based on different levels of aggregation. At its simplest this was state totals from county records. The crop file also aggregates county data too small to reported at the county level into districts. To further confuse matters those districts can be aggregated into a district total. These categories emerged in the extended SAS analysis. The documentation provided with the file does not explain these categories. But other publications from the USDA, and the print versions of the data do explain how the aggregates are organized.

Another SAS analysis, and reading of the documentation, reveals whether there is any potential for reducing the size of the file, e.g., is there any duplicated information that can be condensed into auxiliary tables? SAS descriptive statistics can also 'spot' unreadable data generated or masked by computer static.

The data was also examined to see if there is any potential for enhancements for users. For example, can labels from the codebook be linked directly to the raw data, especially labels such as units of measure and footnotes?

We also encountered problems with the uniqueness of record identifiers. SAS frequency tables, SAS cross-tabulations, and extended SAS analysis revealed multiple observations for a single identifier. We then discovered that in the crops dataset the record identifier field is not unique, it must be paired with the date record punched field.

The statistician spent a considerable amount of time reading background materials on the statistical gathering activities of the USDA and running SAS analyses to gain a sense of how the data were generated: the relationships between observations (especially the different levels of aggregation), and the relationships between individual fields (i.e., yield, production, area).

Quality control

With that background information gathered, the next step in data preparation was to 'clean' the file to ensure proper functioning of the database software so that a search would furnish the desired data. We had to ensure data integrity, and identify potential problems for users.

The first issue in quality control is that of "garbage in, garbage out" or, errors in the data. Examples of such errors are inadmissible values: unlisted codes, numeric values out of range, missing values, unacceptable combinations (FIPS codes for a state county combination that does not exist.) Most errors of this type can be isolated using SAS descriptive statistics, SAS frequency tables, and SAS cross-tabulations.

The data were checked for external consistency through comparison to other data files using extended SAS analysis and printed versions of the data.

They were also checked for internal consistency. Checks were run across fields: *does yield=production*area?; does forage have "not applicable" in production field?* Additional checks were run within fields: do the aggregate level records equal the sum of component values (e.g. *do county production figures add up to the State total?*) do requisite aggregate records exist? (e.g., *do "all" corn record exists if corn subcategory record exists?*) These checks were done using extended SAS analysis.

SAS descriptive statistics, extended SAS analysis (moving average analysis) were used to identify outliers, values wildly out of range compared to other values, e.g., milk production for a particular county being an order of ten more than any other year in the series. Those outliers were identified as probable typos during data entry.

The second issue in quality control is "garbage in, gospel out", inappropriate use of the data. Data coming from a computer has a certain authority not always commensurate with its quality. We could check for certain things that would 'damage' the usefulness of the subsets, such as missing records, using SAS frequency tables, cross-tabulations, and extended analysis, but could not guarantee we had identified all the possible flaws in the data.

Otherwise there was very little we could do to forestall misuse of the data. We consulted the Department of Agriculture publications dealing with how the data were collected and/or derived.

But we decided it was not within the purview of this project to attempt to recommend ways to use the data for statistical analysis. We could not tell users the adjustments required, and the limitations of the data for a particular statistical inference. Instead, we relied on giving users the citation information they would need to locate the detailed data collection/derivation documentation from the publications of the USDA.

Mounting the data into INFeRS

The data were downloaded from tape into Albert. Data modifications were made using the most efficient mix of CMS commands, SAS, C, and Informix. After the dataset was loaded into Informix, a systematic battery of sample extractions was run against INFeRS to verify that the system was producing the correct subsets. Those subsets were checked for agreement with a SAS extraction run against the raw dataset.

In the case of the 'super-checked' Crop Estimates file, the statistical consultant learned (through conversations with the U.S. Department of Agriculture's National Agricultural Statistics Service) that state agencies are responsible for corrections to their data. As a result he actually wrote to all the states producing datasets with possible errors requesting they confirm any of our corrections made to ensure proper functioning of the database.

Recommended Standard Operating Procedure in the Future

There are two categories of procedures to consider: level of desired quality control and level of data modification allowed.

QUALITY CONTROL

Curative: Do only what is required to download the data and mount it into INFeRS. Data problems are resolved subsequently as they arise. Set up the tape, move the data, verify they are the right data, identify the characteristics/problems the programmer needs to know to facilitate proper design of database and interface and to facilitate proper mounting, and identify characteristics/problems the user needs to know.

Preventive: Do sufficient cleaning to ensure INFeRS will perform efficiently and smoothly with the dataset. Somehow condense duplicate information, correct unreadable data, and resolve any problems with the uniqueness of record identifiers.

Seatbelt: Tailor the file for users by identifying/correcting potential problems or errors detectable in the data to facilitate extraction of the desired data and subsequent applications, and provide limited documentation on characteristics of the file, such as identifying missing values and 'non zero' zeros.

Airbag: Provide the user with sufficient information about the dataset to ensure proper application of the data, particularly about their statistical properties. Document aggregations, how the data are generated, and their statistical properties.

Nirvana: Ensure data is as free of error as humanly and computerly possible. Do everything to clean the data and document it for the user.

LEVEL OF DATA MODIFICATION ALLOWED

Prime Directive (or WYSIWYG) Strategy: The integrity of the original data is strictly maintained, even if it compromises the efficiency of the database software. Warnings of potential errors are generated with exported data.

Unsecured Limited Intervention Strategy: Data modifications are made as needed to ensure the smooth functioning of the database software.

Secured Limited Intervention Strategy: Same as "unsecured" except modifications are submitted to the data source for verification.

Missionary Strategy: Make any changes we decide are likely to improve the quality of the data; again with "secured" and "unsecured" versions.

RECOMMENDATION

The most realistic, yet responsible procedure would be the Seatbelt strategy (correct the data to the level needed to ensure reasonable data extractions), with Secured Limited Intervention (verifying data modifications).

Data Coding Categorization

One by-product of data preparation was our categorization of data coding. The interface was designed to accommodate the types identified. Adding new files into the system would involve identifying how that file's data coding scheme fits into our previously identified categories. New design would only occur if new categories of data coding were encountered.

The seven major categories we encountered were:

— Coded — simple: a small number of codes, enclusively alpha, or numeric, or a combination of both.

— large: an extensive number of possible codes, enclusively alpha or numeric, or a combination of both.

— Mixed coded, numeric: values coded up to, or below, a threshold number.

— Dependant coded: Field B applies only when Field A has a certain value.

— Special Fields: the definition of Field B varies, depending on what is coded in Field A.

— Hierarchical: Field A is further subdivided into Field B, etc.

— Numeric: range: numeric values for the field all fall within a certain range.
unlimited: numeric values for the field have no range.

— Related fields: Field A, B, and C are all defined the same, allowing up to three equally valid descriptions of the sample.

System Creation

Introduction

INFeRS was designed to supplement existing numeric data retrieval methods for the research community. In doing so, we addressed three technical objectives: 1. develop a structured, interactive search and retrieval environment capable of supporting unstructured data requests, 2. explore the issues involved with providing computerized access to large statistical data sets versus bibliographic data sets, and 3. completely remove the end-user from the need to know how the data is stored and what is necessary to access it. This concept contrasts the traditional Data Archives approach to data retrieval.

The Data Archives model requires that a custom computer program be written for each new data request. This in turn requires that someone be familiar enough with the programming tool to be used, as well as specifics about the physical layout of the data, to be able to identify appropriate fields and construct search conditions. Data sets accessed under this model are traditionally stored off-line (magnetic tape storage) and require sequential access when programs are run against them (data set read from beginning to end, regardless of the amount of data requested). INFeRS, on the other hand, was designed to be an on-line (disk storage), interactive (direct access to requested data) tool to serve the same data retrieval purpose. Specifically, to allow a user to specify variables, search conditions and output formats of interest, and to execute a search and download the results to a workstation for analysis without programming.

Computer hardware and software selections for the project were the result of extensive evaluation processes by the library based on specific needs of the project. Relational database technology was chosen as the data management and application development tool, and the data and software reside on a Hewlett-Packard 9000 825 server running HP-UX (UNIX System 5). Access to the data is gained from researcher workstations over the campus network, a TCP/IP-based Proteon network. Researchers without access to the campus network can use RS-232 or telecommunications devices. The workstations in common use by the research population are Macintosh, IBM compatibles and UNIX workstations. The C programming language was also used to support areas of the system not best handled within the Informix programming environment.

Project Phases

The creation of INFeRS was divided into three major development phases. The phases were each designed to emphasize particular tasks, with post-Phase III work intended to be the addition of new data files into an established system. The identification of system specifications was a combination of initial project objectives, end-user interviews and extensive data analysis by the INFeRS project team of the data sets previously selected for inclusion in the project. Once identified, the requirements were then analyzed and parceled into project stages to create the development phases described below.

Phase I consisted of the development of basic system functionality: designing a relational database around an existing data set of predominantly coded and numerical data, identifying and building supporting data structures necessary for the program to run as intended (variable definitions, program lookup tables, etc.), building an interface to allow a user to define search conditions on various types of data, developing the program logic necessary to construct an executable search statement based on any volume and combination of user input and providing supplemental output to document the search. Emphasis at this stage was on creating a system that "worked"—i.e., enabled all search components to be specified by the user, executed the search, and output the results—with the data access 'know-how' to be transparent to the user.

Based on end-user evaluations of the Phase I system and increased knowledge of the development environment, Phase II introduced a new set of tasks to build on the accomplishments of Phase I, particularly, by accommodating a more complex data set, incorporating more sophisticated search definition and display features, allowing a wider range of output possibilities, designing a more appealing and user-friendly interface and dramatically improving search execution response time.

The purpose of Phase III was to take the end-product of Phase II and reapply it to the Phase I file. The intent was to maintain the existing Phase II interface and functionality as much as possible, making major modifications or enhancements only where necessary to accommodate unique elements of the new file. This effort explored the similarities and differences across the files, taking advantage of the similarities where possible. Phase III did require the addition of a few "new" components to apply existing concepts to new situations.

System specifications, design and development

The development of INFeRS to its current state is best described by outlining the characteristics of each phase of the system including the discoveries encountered. Although the underlying objective of the system never changed across phases, not all tasks could be accomplished at once much less be presented to the user in the most attractive or logical way the first time around. Therefore, as outlined in the introduction, it was necessary to tackle basic system functionality first, then entertain more attractive and sophisticated methods of performing the basic functions as the phases progressed. Basic system functionality consisted of:

—Selecting a database of interest.

—Making various choices to develop a search strategy.Based on a series of hierarchical menus and windows.

—Selecting additional non-filter variables as output-only (all filter fields are automatically tagged for output). This feature was not implemented until Phase II.

—Selecting an output format of choice and specifying a meaningful file name (for file output, only).

—Executing the search.

—Selecting a file transfer method and initiating the transfer of data to an end-user workstation (for file output only). This feature was not implemented until Phase II.

—Loading search results into another software package for analysis and interpretation (for file output, only). This task to be accomplished outside the INFeRS system by the end-user.

Phase I objectives

DATA—provide access to the USDA County Crops Estimates, a small file in terms of total variables, by designing a database to support it, and all other lookup, validation and program tables necessary to run the system around it.

INTERFACE—make use of Informix programming language tools for quick development and prototyping of ideas by removing the need for extensive programming effort to implement initial program flow.

—implement a simple menu hierarchy that logically supported the progression of activities necessary to build and execute a search.

—provide context-sensitive help throughout the program.

SEARCH DEVELOPMENT AND EXECUTION—create the modules necessary for a user to select fields for search conditions. Search conditions could be based on both coded and numeric fields.

—require mandatory use of particular fields during search strategy definition to monitor major elements of the search. This was necessary to avoid potentially extracting the entire file or very large chunks of it.

—design and build the program logic necessary to construct an executable SQL statement based on any volume and combination of user criteria.

OUTPUT—make use of the C programming language to handle all file management and data output routines.

—define and provide basic screen display and output formats to facilitate easy integration of search results into popular software packages (SAS, dBASE, Lotus, etc.).

—automatically generate full-record output, removing the need to implement the logic required to handle variable length output files.

—design and build the program logic necessary to create a custom information file to accompany each output file defining the specifics of the search that produced the subset.

Phase II objectives

DATA—provide access to a more complex file, the SCS National Resources Inventory. It is a significantly larger file in terms of total variables, with data dependencies and 3 levels of geography.

INTERFACE—invest more programming effort to develop a more attractive and user-friendly menuing environment while still maintaining the basic flow established in Phase I. The final structure consisted of a series of drop-down menus allowing horizontal and vertical movement between and within the various menus. The technique used to develop this feature was carried over into all browse and display windows. This not only enhanced the look and feel of the overall interface, but also provided vast control over key-initiated actions not possible in Phase I under the 4GL approach to menus and screen control.

—invest more programming effort to perfect search definition modules, allowing much more flexibility for all browsing and data entry environments.

—pay more attention to general screen presentation, continuity of key use and system flow.

SEARCH DEVELOPMENT AND EXECUTION—develop two new methods, referred to as indexes, of locating database fields of interest. From either index, any field can be selected for output only, or used as a filter field, in which case it is automatically selected for output.

The "Subject Index" was designed to allow access to a field based on any qualifying keyword or subject term. Identifying and locating a subject term having anything to do with a field allows access to the field, and is targeted toward people less familiar with the data. This feature also permits the user to use the alphabetic keys for faster movement within the index window by using them to quickly move forward and backward.

The "Field Index" approach has only one access point for each data field, but logically groups the fields into a subset of category terms, providing easier access to a field for people more familiar with the data set.

When a coded field is selected from within one of the indexes, a lookup is performed to locate all valid values for the field and displays the first "page" of them in a selection window with the first entry highlighted. Any previously selected entries for the field are tagged and selections are made or removed by highlighting the item and pressing the return key.

When a numeric field is selected, numeric expressions can be built. To build a numeric expression, a selection window first allows choosing from a list of available operators (>, <, =, <=, >=). Once an operator is chosen, a data entry window displays either 1 or 2 data entry points, depending on the operator choice. Error checking takes place to check for "reasonableness" of the expression and to identify valid ranges and exceptions.

—require a minimum condition based on one of 3 levels of geography. To make it easier to locate geographies of interest, incorporate a "lookup" feature that permits substring searching on entered text versus straight hierarchical browsing.

—create a "Show Strategy" feature to display the current status of a search at any time for geography and non-geography conditions and output fields.

—design and develop better internal query construction and execution logic to improve response time.

—enhance the Phase I help module by converting it to a hierarchical menu system for displaying general system use and movement assistance as well as codebook information for fields.

—allow the ability to terminate a search mid-execution.

OUTPUT

—permit user-definition of an output variable list versus full-record output.

—allow file transfer capabilities.

—incorporate better output format possibilities and make enhancements to the information file generated by the search.

File formats included: delimited (Tab, pipe, Basic) and non-delimited formats. Information file contents were: all search parameters (field names and codebook information); other descriptive information (disclaimer, notes, warnings); and total number of records extracted.

Phase III

Phase III took the end-product of Phase II and applied it to the USDA County Crops Estimates again (this file was initially loaded in Phase I). The Phase II logic and interface was, for the most part, maintained. Modifications in the form of changes to existing modules, or the creation of new ones were developed only to accommodate aspects of the Crops data that did not apply to the National Resources Inventory. Examples of necessary modifications include:

—units-of-measure differences between commodities requiring that the user be notified when defining relational expressions on numeric fields to avoid "nonsense" searches. This was accomplished by incorporating another lookup table to track the units of measure differences across commodities and the various numeric fields.

—an enhancement to the Show Strategy module to display the interpretation of all defined relational expressions as a second means of alerting the user to potentially "nonsense" searches.

—adjustments to the query construction and execution logic to convert the Phase II logic (NRI) to handle the Phase III data (Crops).

Query Construction and Execution

An interesting trial-and-error aspect of the system was designing the proper programming logic to build and execute an SQL statement from scratch. The program must build and execute a search statement based on user input, but Informix "takes over" from there because of a feature unique to relational technology, the query optimizer. Unlike general programming languages, the query optimizer accepts a request in the form of an SQL statement, analyzes it, then decides the best way to tackle the problem to locate the answer. These access decisions are based on several things Informix knows about the data and the way the request was made. Therefore, writing a program to construct a statement that returns the expected results was the first challenge. Influencing the optimizer to make specific decisions about how to perform the search was the next challenge.

To build an SQL statement from scratch, the program must process and piece together all the

conditions that have been specified and stored during the session (individual numbers or ranges of numbers for relational expressions or alphanumeric codes for matching, for example). To do this, the program logic must handle situations such as:
—determining which variable the current condition belongs to
—identifying the data type of the variable to build either a relational expression or selection list, whichever is appropriate
—deciding if this is the first condition encountered for this variable
—if so, complete the previous variable expression if there was one, and start a new expression based on the current field and condition, etc.
—if not, concatenate the current condition to the statement according to the data type

As the user creates search conditions, the conditions are held in a temporary storage area (in a temporary table) to allow the flexibility of modifying the search at any time. This makes it very easy to process them when it is time to execute the search. Each stored condition (code, number, range of numbers) also has an identifier to link it to the appropriate variable in the database. This identifier is used to group the conditions by field for processing to construct the search statement, identify the "first" and "last" conditions for any given field to "start" and "end" each new variable expression, and output search-specific documentation to an information file for future reference.

Although there were many ways to make the program assemble a correct SQL statement, our ultimate goal was to identify a method that was both fast and efficient. The speed issue is obvious—because a user is working with the system interactively, the expectation of immediate response time is much higher than if a batch program is being submitted. An inefficiently constructed search (in terms of complexity and/or overall length) can have a significant impact on response time, especially for large amounts of data.

The Phase I solution produced a long and complex statement which required unnecessary work for Informix. Indexed fields were referenced first to ensure index use, but Informix didn't know "where to stop looking" for key fields within the statement. This resulted in unnecessary scanning, therefore longer searches and, ultimately, complex searches taking a long time to run. Output field lists were not an issue at this stage because the basic objective was to output the full record.

The Phase II solution was significantly different from the Phase I solution. Here we concentrated on devising a way to minimize the length of the overall statement as well as cater to efficient index use. Because geography was a mandatory condition variable, it was known that at least one geography selection would exist for every run. We used this to our advantage by excluding all references to geography from the filter clause. Instead, geography selections were drawn on one by one from their storage area as the search executed.

We accomplished this by using a popular relational technology concept called "joining." When a search is executed, two activities take place very quickly. The join condition instantly identifies rows meeting the specified geography by applying the stored selections, one by one, to the index. The optimizer then imposes the remaining filter conditions, if any, on only the rows that satisfy the join. Rows that satisfy the join and pass the filter are added to the output subset.

Once the join logic was functional for geography, it was possible to incorporate other indexed fields into it (frequently used, non-mandatory fields). Because the additional join fields were optional conditions, the program had to decide if any of them were taking part in the current search. If so, they had to be folded into the logic so that all unique combinations of variables taking part in the join were generated and accounted for.

Another specification of Phase II query construction was the ability to accommodate variable output field lists. After some experimentation, we concluded that the most efficient way to accomplish this task was to retrieve the entire record then impose the selected field list as part of output processing.

Phase III converted the Phase II query construction logic to apply to the Crop Estimates file. Adjustments had to be made to account for the different file, however the transition went rather smoothly and should be applicable to other files.

Database design and construction

There were many interesting and challenging aspects of designing a database to support INFeRS. As it turned out, we developed separate databases for each file loaded into the system, carrying identical tables of data across databases (state/county lookup tables and system messages, for example) and dealing with unique design situations as each new data file was incorporated into the system.

Normally, a database is designed to computerize an existing function, series of activities or information needs (payroll accounting or course registrations, for example). In these situations a detailed analysis process will identify the activities and data processing needs for a solution in the form of system specifications. These specifications can then be used to identify the entities and entity relationships that the database must support to achieve the system requirements. The outcome of this process is a definition of the raw data necessary to be maintained in the database and processed by the system for the system to function as required. INFeRS forced us to start from the other end, with existing data and, once system specifications were identified, design a database around the data and develop processing requirements to best meet the specifications.

Database design issues included the following items:

—It was necessary to improvise on some database design rules due to the constraints that existed when dealing with pre-defined data (existing coding schemes, data integrity issues, etc.). Some examples of the data problems encountered were: non-uniqueness in the data where uniqueness was expected, erroneous data and variable dependencies causing multiple data types to exist within "one" variable.

—We identified appropriate data types for variables based on analysis of the raw data and its supporting documentation.

Three types of data were supported by the database and were necessary for the creation, execution, and documentation of searches:

—base data set (raw data)
—supporting data we had to locate and load (definitions and validation tables)
—program data we had to identify and create (help text, menu text, output documentation and system messages).

Beginning with Phase II, extensive time was spent on the identification of indexing strategies that would provide uniqueness as well as facilitate searches on frequently-used or "common" variables. It was not practical to index every field, but it was very advantageous to identify strategic fields and field combinations that would influence response time for a majority of searches. For example, geography was determined to be a significant filter item, therefore geography fields were always incorporated into an indexing strategy (three separate levels of geography existed, and were made searchable, in the National Resources Inventory).

Also, in the effort to enhance response time, temporary tables were designed to be created and used for the duration of an INFeRS session to hold user selections (filter conditions) and assist in executing the search.

Data Problem Resolution

One element of the data analysis effort was the identification of potential 'problems' with the data that could hinder user access. Once identified, we had to decide how to deal with each situation to minimize the adverse impact, such as the potential of developing 'nonsense' searches, or missing data altogether.

We identified two basic methods of handling such situations. The first was additional programming to account for the problem without user intervention. The second was additional system functionality to make the user aware of the situation and allow them to select the appropriate action.

Internal solution

We applied this invisible-to-the-user approach to any area where it was possible to put the system 'on the lookout' for certain situations and to make necessary adjustments not requiring further user intervention or decision-making. The necessary action is static across searches, so it can be accounted for within the program. By developing additional program code to check for an instance and take the appropriate action these situations are handled entirely transparent to the user.

For example, we applied this internal approach to searches based on the Conservation Practise 1, 2 and 3 fields within the National Resources Inventory. These three fields are designed to hold a maximum of three conservation practices for each record, with no hierarchy intended as to what may be in each field. Unless we made it possible to define conditions on these fields in one central place it would be necessary for the user to define the same set of conditions three times, once for each field. Requiring the user to do this is cumbersome, confusing, and prone to error.

To forestall that, we collapsed the Conservation Practise 1, 2 and 3 fields into one field: Conservation Practise. Any definition in this field is applied to all three actual database fields. Filter conditions based on Conservation Practise are expanded out and run against all three fields when the search is executed. Also, selecting Conservation Practise for 'output only' expands out to include all three fields when the search is assembled for output.

External Solution

We implemented this user-beware approach in places where a warning message was warranted, where it was possible to generate a "nonsense" search, or if a decision needed to be made before continuing. These problems are handled within the interface, not the program, so that the user is aware of the situation and can decide whether to continue as is, or make any necessary adjustments.

We applied the external approach in an instance of field dependency that could not be handled without user input. The Diameter at Breast Height (DBH) and Basal Area/Stem Count fields within the National Resources Inventory have a relationship where the DBH value for any given sample point/record dictates whether the contents of Basal Area/Stem Count will be numeric (Basal Area), or coded (Stem Count). Because each field may or may not be included in a search strategy, the user must know the status of one field if the other is about to be used as a condition field.

We developed a series of message screens that are invoked when either of these fields is selected. The screens explain the relationship between the fields and describe how to develop a "sensible" search. The screens then allow the user to select either Basal Area or Stem Count so that relational expressions can be defined when searching on the former, while coded value selections can be made when searching on the latter. Although the program does not check for or enforce "sensible" searching on combinations of these two fields, it does impose enough information to allow the user to make a proper entry.

Technical Conclusion

It is important to know the data as well as possible prior to database design in order to reduce the need to undo, rethink, redesign, and rebuild, and in order to develop indexing and query execution strategies to enhance search response time.

Although it is difficult to anticipate issues relating to data provided by an outside source, upfront data analysis performed by someone with statistical expertise is time well spent. A database change identified prior to database design can be accommodated better than a database change identified during program development. This is because the impact of a change is magnified when existing program components are involved and must be modified, and should be avoided when possible.

In a relational database environment, query response time can be significantly improved by the use of indexes. This only applies, however, if the indexes are created and used effectively. If 'common' search fields can be identified and incorporated into indexes, program logic can be developed to ensure their use whenever appropriate. Although it is unlikely that all fields can be accommodated in an indexing scheme, a significant 'first cut' at the data can be made if significant fields are identified, indexed and enforced during query development.

Relational database technology is appropriate for managing and querying large statistical data sets. Developing an interface to define and execute the queries, however, tends to impose searching constraints that do not exist with free-form SQL or custom programs. It is important, therefore, to recognize and address the data needs that can be met by providing interactive access to a data set as well as those that may require individual attention.

Program design decisions concerning when to 'shield' the user from data concerns versus 'warn' the user of their existence can be very challenging. A programmer familiar with the data can choose to allow for, or omit, certain situations when writing a custom program to perform a search. An INFeRS user on the other hand, is limited to the decision opportunities that the application is designed to provide. The challenge, therefore, lies in deciding when a decision is best handled by the individual user (and what the list of choices must be) and when it is appropriate for the program to be developed to handle the situation internally.

Significant time can always be invested in designing and implementing a more attractive and user-friendly interface. This should be balanced, however, with investments in providing improved functionality and overall query response time as well as general system maintenance and support.

Because a typical output extract from an INFeRS search consists of unlabelled rows of data, it is vital that every search be well-documented to ensure that the data is understood and loaded properly into a data analysis tool. At a minimum, the basic codebook information should be provided such as: search conditions, field names, field definitions, valid values, and output column positions. In addition, we found it useful to include supplemental text, and references to other sources.

Developing a controlled searching environment can be very helpful in the overall development of an interactive system. The end result of providing selection windows and reducing the need for data entry wherever possible is better end-user searching.

Subject Indexing

One of the project objectives in developing INFeRS was to provide users with sufficient information on the data resources residing in INFeRS to enable them to effectively formulate searches specific to their needs. Library technical services departments have traditionally been responsible for the acquisition of materials and their organization for use. A critical part of the organization process is the analysis of materials to identify the subject content and to assign appropriate subject access points. The INFeRS project required that the electronic numeric data files be subjected to a similar process. The goal was to describe the basic subject content of the INFeRS datafiles as well as to provide users with a list of all the variables represented in the files. This was accomplished by constructing uniform keyword indexes for both the USDA Crops Estimates—County File and the National Resources Inventory database.

Indexing all significant keywords as subject access points was necessary to make the system user friendly for the "naive" user unfamiliar with the content of the databases. This would enable a broad spectrum of users, both local and remote, with abilities and needs of varying degrees of sophistication, to successfully utilize the system. The keyword indexes guide the user through lists of possible search parameters or variables from which to develop search conditions. The selection of variables for searching is aided by listing them in a logical fashion, gathering related variables under broader concepts to alert the user to all possible search conditions related to their topic.

The indexing was accomplished through a detailed analysis of the codebooks for the USDA Crops Estimates and NRI databases. Key subject words were identified from the field or variable names, the field descriptions and the code descriptions of the possible values specified for each data field. The list of keywords was augmented by identifying broader keyword concepts under which to collect related fields. Additionally cross references were provided to lead the user from variant terms (e.g. inverted or modified terms) to the selectable term for a variable. This process resulted in two separate indexes for USDA Crops Estimates and the NRI. The Crops Estimates keyword index (constructed in about 5 hours) consists of 57 terms or phrases and 6 cross references describing 10 data fields. The NRI keyword index (constructed in approximately 29 hours of work) consists of 298 terms or phrases and 57 cross references describing 69 data fields. The basic arrangement of the indexes is alphabetical by keyword, keyword concept and cross reference term. Any given data field may appear numerous times, depending on how many subject areas it relates to. Not including help text, the index runs for 66 selection screens. Movement through these screens is assisted by the capability to "jump" alphabetically into a section of the index by entering the initial letter of a desired term. Formatting of the keyword indexes for loading into a keyword table was determined in concert with the project programmer and the interface designer. This formatting provided the system with the information necessary to lead users through a logical progression of selection windows in formulating their searches.

In its final form on the INFeRS system the keyword index is referred to as the Subject Index. A second index, the Field Index, provides users with an alternate way of formulating search conditions. The Field Index groups related fields into five broad categories, plus the SOILS-5 Identification block. Each non-geography field is indexed only once under a single category. Selection of a category provides the user with a list of field names alone for further selection. This index was designed primarily for the experienced user familiar with the content or terminology of the particular datafile. It provides a much quicker means of access since the user is confronted with only one or two selection screens for any one category of fields. The Field

Index obviated the need for what had been termed a Sequential Index. This index would have listed all fields, each only once, in the order in which they appeared in the codebook. It was originally felt that anyone familiar with the documentation for the NRI datafile might prefer this to a categorized approach. Testing of the three indexes indicated that the keyword and category approaches were the most useful and the Sequential Index was dropped from the system.

The INFeRS project was designed to address a number of questions regarding the role of libraries in providing computerized access to numeric data. Several of these questions related to determining the demands non-traditional electronic information resources, such as numeric data files, place on libraries in terms of time, resources and staff skills. There is a clear need to define the role of technical services in mainstreaming electronic information resources into traditional library collections. Aside from the need to identify and acquire these resources, the technical services component of the project focused on the processing and analysis of numeric data files for their subject content in order to facilitate their use.

Although subject analysis is a normal component of the cataloging process, the development of an interactive online system to access a non-bibliographic form of information did necessitate looking at the process in a new light. Traditional subject cataloging involves the use of an indexing language, a controlled vocabulary of accepted subject terms. Although an established vocabulary such as the Library of Congress Subject Headings (LCSH) could be used, it would likely be cumbersome, less specific and ultimately unsatisfactory when applied to a concise, well-defined collection of data. Given the more specific subject context of numerical data files, users are probably better served with indexing based on the terminology of the files' codebooks. The technical services member of the INFeRS team (a monograph cataloger), therefore, had to create the vocabulary to be used in the indexes. The index language had to conform to that used in the variable or field names, the field descriptions, and the codebooks in general. However, there was a perceived need to also maintain some consistency with the indexing language users normally find in library catalogs and indexes, namely LCSH. The syntax of LCSH is complex, utilizing both simple terms and phrases, inversions, parenthetical qualifiers and extensive subdivisions. The logic of LCSH is not necessarily consistent, obvious or intuitive, but it should be taken into consideration in striving to provide users with some degree of uniformity of access to library-based information systems. The INFeRS subject indexes consist of keywords from the code books rather than LCSH terminology, but they do borrow elements from the LCSH syntax. Parenthetical qualifiers and inverted headings (as headings or cross references) were utilized, as were collective terms (keyword concepts). Since catalogers are used to working with an existing indexing language rather than having to create their own, the indexing of the numeric data files proved to be an interesting exercise. It also indicated an area worthy of more study.

Subject Indexing Conclusions

While subject analysis skills are already resident in cataloging staffs, programming skills generally are not. Few catalogers have experience with the type of relational database management package required by this type of project. In the absence of this capability, technical services staff should be able to work with programmers in formulating standard procedures and generic instructions providing access to non-bibliographic datafiles. In the future it is not inconceivable that catalogers will have developed all the skills necessary to routinely integrate newly acquired datafiles into the established access structure encompassing all library resources.

File Transfer Module Development.

INFeRS differs in two ways from most remotely-accessible information systems that are currently available to the Cornell community. First, INFeRS outputs the information retrieved to a set of files: one for data, one for codebook information. Other remote information systems generally output any information retrieved to the user's display. Secondly, the quantity of data retrieved from INFeRS is relatively much larger than that retrieved from most other information retrieval systems. Based on our experience with the working Phase II prototype, most searches of the National Resources Inventory numeric datafile will retrieve between 500 and 3000 records, which would be of a size on the order of hundreds of kilobytes (KB). Given that a full page of single-spaced text is approximately 4 KB in size, hundreds of kilobytes is roughly equivalent to between 25 and 250 pages. A user searching a bibliographic index and abstracts database such as ERIC typically retrieves between 30 and 60 records, which would be between 60 and 120 KB in size or roughly 15 to 30 pages.

Text capture has been the traditional method of downloading data from currently available information systems. With the text capture method, a user activates a feature of the communications software running on their local computer that concurrently "captures" the text that appears on their screen and prints it to a printer attached to the user's computer or saves it to a file on a diskette or hard disk on the user's computer. The speed of the text capture method is limited to the speed at which the text scrolls by the user's display. Text capture can be particularly slow if the user is accessing the remote information system via a modem. For downloading data that is small in size (e.g. less than 100 KB), text capture is generally satisfactory. This method is not practical or reasonable for moving files of the size typically extracted using INFeRS. Consequently, we explored the use of formal file transfer protocols as a solution to the problem of getting the information retrieved using INFeRS back to the user's computer.

A *file transfer protocol* refers to a standardized mechanism by which data can be transferred as a file rather than character by character as in the text capture method. Two constraints had to be satisfied in our selection of a file transfer protocol for use with INFeRS. The file transfer protocol had to be supported in (1) the computer networking environments at Cornell and (2) the computing environment of our user population.

Cornell University maintains two campus computer networks. The campus Sytek network is a high-speed serial communications network. In functionality, it is much like a very fast modem.

The campus Internet network is a much higher speed communications network based on TCP/IP network protocols. The campus Internet links Cornell with the larger national Internet network. With respect to these two different networks, several file transfer protocols exist for use in either networking environments. Kermit, X-MODEM, Y-MODEM, and Z-MODEM are the most widely known file transfer protocols developed for a serial communications-based networking environment like the campus Sytek network. For TCP/IP-based computer networks, two file transfer protocols exist called tftp and ftp (*tftp* is an acronym for trivial file transfer protocol; *ftp* is short for file transfer protocol).

During a publicity campaign to inform Mann Library's constituents of the INFeRS system, a questionnaire was distributed to survey the computing environment of this group. Analysis of this survey revealed that the pool of potential INFeRS users made use of a variety of communications software to connect to remote computers via both of Cornell's computer networks. In most cases, a user had access to only one of the two campus networks from their office or department. Therefore, INFeRS would need to support file transfers using a protocol supported by each of the two campus networks. The most heavily used communications programs also had the capability to do file transfers using either of the Kermit or ftp file transfer protocols. Consequently, the decision was made to design an automated file transfer feature into INFeRS which supported the apparent de facto standard protocols, Kermit and ftp.

Foreseeing a need for a file transfer feature in other information systems under development at Mann, a general-purpose file transfer utility program was implemented, independent of INFeRS. As an independent program, the file transfer utility could be used by future information systems developed at Mann as well as INFeRS.

To implement the automated file transfer feature, a second programmer on the Mann Library staff, Tim Lynch, was enlisted in the project. He was responsible for implementing the portions of the file transfer utility that managed the Kermit and ftp file transfers. User interface design and programming for the file transfer utility was done by the interface designer. The interface screen of the file transfer utility displayed the files to be transferred, their size, and an estimate of the time needed for transfer. To incorporate the file transfer feature in INFeRS, the project programmer designed and implemented two screens, from which a user selected the set of data and codebook files to be transferred and the file transfer protocol to be used. Once the selected files and transfer protocol was obtained from the INFeRS user, the information was passed to the file transfer utility, which invoked the appropriate file transfer program and initiated the transfer process.

Interface Design

The role of the user interface designer in Phase II development of INFeRS

In July 1990, William Garrison joined the staff of Mann Library as a user interface designer and systems analyst. Shortly thereafter, he became a member of the Title IID INFeRS project team. At this time, the team had completed development of the Phase I database application and was beginning design of the Phase II application. During Phase I of the project, well-defined roles had evolved for the various team members. As expected when a new member joins an established team, there was a period of re-orientation while the interface designer was integrated into the team. While it was clear that the interface designer's role would be to contribute to the user interface development of the Phase II application, it was less clear at first how this would be done within the context of the project team.

The addition of a user interface designer to the team had the greatest impact on the role of the project programmer. In developing the Phase I application, the project programmer had been primarily responsible for the design and programming of the interface as well as the underlying database application. With the addition of an interface designer, the programmer's role in Phase II of the project would be redefined to have less responsibility for interface development. In view of the fact that the interface designer also had programming experience with a number of high-level languages, dividing the design and programming of the user interface and the database application between the interface designer and the project programmer was considered.

Associated with this approach was a high startup cost in terms of project time and effort. The interface designer had no experience with Informix, therefore some time would be needed to learn the interface development language. The most prohibitive factor of this approach was that the program design to be used for the Phase II application could not accommodate separate development by two programmers. The project team intended to use the design developed during Phase I as the basis of the Phase II application. Under the development of a single programmer, the Phase I program design had evolved to a point where the programming for the user interface and the database was highly intertwined and could not be easily separated. Given the intertwined nature of the interface and database components, it was impractical for two people to work individually on those components because both programmers would need to be mindful of interdependencies in the design. Redesigning the program to make the user interface and database components more independent would have required considerable effort, and a significant amount of time out of the project schedule.

Given the concern of the project team to remain on schedule, the decision was made for the project programmer to continue as the primary programmer of both the interface and database components of the Phase II application. The interface designer would provide consulting to the team on user interface issues and work closely with the project programmer to review the design and implementation of interface screens.

The process of interface design for the Phase II application

The user interface design process started in the weekly meeting of the project team. Interface design issues were raised and discussed so that input could be gathered from the diverse viewpoints represented in the group. Project team discussions resulted in general ideas for screen design. These ideas would be refined in a later session between the project leader, programmer, and interface designer.

Ordinarily, screen design ideas were immediately implemented by the project programmer in the Informix development environment. Once the initial user interface programming was complete, the interface designer reviewed the screens with the programmer and proposed suggestions for modifications. The implemented screen design was also reviewed by the full project team.

When scheduling was not critical, screen design ideas were prototyped by the interface designer as paper mockups or partially animated simulations using software on a Macintosh computer, rather than being directly implemented by the programmer. The screen prototypes were reviewed with the project programmer to identify technical limitations. These prototypes were also reviewed during meetings of the full project team and were then modified to reflect any changes in the design. The modified screen designs prototypes were then used by the project programmer as "blueprints" for implementation in the Informix development environment.

Influences on the user interface design of the Phase II application.

The process of user interface design is one of continual tradeoffs. In a perfect world, user needs would take top priority in all aspects of the design process. In reality, constraints are placed on the design of any interactive computer system from many fronts such as the end-users, the development environment, the operating environment, or the organization. As a rule, interface design decisions require tradeoffs between the various constraints in order to produce a system that is both usable and practical to develop. For the Phase II version of INFeRS, the major constraints on user interface design were the data format and relationships between variables in the datafiles, the interface development software, and the project team's approach of direct implementation of preliminary interface designs.

Data format and relationships between variables in the National Resources Inventory (NRI) datafile were a major influence on the user interface design. N early half of the variables in the NRI datafile are categorical variables, i.e., each value represents an option within some category. The values of these categorical variables are numerically coded. For example, the code for Illinois in the variable *State* is the number 17. Without a codebook, it would be difficult to interpret what the coded values of a variable represent. To make these categorical variables intelligible to a user without the codebook, a design decision was made that the corresponding labels would be displayed in addition to the coded values whenever a categorical variable was presented to the user. The NRI datafile also contains many variables which have dependent relationships with other variables. For these variables to be properly interpreted, the dependent relationships have to be displayed to a user. To facilitate the display of these relationships, the screens for selection and searching of these specific variables required special user interface design. For example, the producers of the NRI datafile used three variables to represent a three-level hierarchical organization of categories for land cover and use. For the logical selection of values for land cover and use to be made by a user, the hierarchical relationship of the variables had to be displayed. As a result, a unique three-layer menu screen was designed.

The interface development facilities that were incorporated into the Informix-4GL database software more often than not constrained rather than facilitated our screen designs. There were various and seemingly arbitrary restrictions to the manner in which text could be positioned and formatted on a screen. For example, three lines of every window of text were always reserved for use by the Informix interface software and several control key combinations had predefined uses which also could not be contradicted. Informix interface development software was used for Phase II development because it provided mechanisms for creating and managing menus and online help that were integrated with the database development software. While the interface development software had the potential advantage of reducing development time, its inflexibility discouraged us from using it. The project programmer discovered a feature of the interface development facility that allowed her to program around its limitations. Consequently, the majority of the Phase II application's user interface was implemented with custom designed programming by the project programmer.

Rather than prototyping screen design ideas, the project programmer directly implemented the ideas in the Informix programming environment. This approach had a stifling effect on the the user interface design process. The project programmer invested a significant amount of development time designing the program structures and code to implement a particular screen design idea. Cosmetic changes to the design, such as shifting of the layout or rewording of text, were relatively quick and easy to make. Major changes in the design, such as reorganization of the screen or changes in keyboard methods used to navigate through a screen, would often require extensive redesigning and reprogramming. Pressure to keep the development on schedule made major modifications prohibitive. Consequently, suggestions for minor modifications could be implemented, whereas suggestions for major modifications were typically shelved in hopes that time would be available later in the schedule to accommodate them. In practice, the development schedule was unrelenting, so that the programmer's initial implementation of screen design ideas eventually became the final screens of the user interface for the Phase II database application.

The role of the user in the design process

Incorporating user feedback in the design process of an interactive computer system is a recommended practice. User evaluation early in the development cycle can be valuable in identifying confusing areas in the system and can lead to the discovery of features that have been overlooked. After implementing the preliminary design of Phase II INFeRS, ten people reviewed and evaluated the system. All reviewers were Cornell faculty and staff. Some had experience with using numeric datafiles in general and the National Resource Inventory datafile in particular. Others had little experience with numeric datafiles but were experienced computer users. One reviewer was a user interface designer from another department at Cornell. Each reviewer executed a series of searches of the NRI datafile that were prepared by the project team to systematically introduce the user to the different parts of the interface. Reviewers were also permitted to formulate their own searches if they desired. Reviewers examined the system in the presence of one or two members of the project team, who solicited and recorded each reviewer's comments and answered any questions that arose.

Upon completion of the evaluation, the comments of the reviewers were compiled and combined with observations made by the project members. Analysis of the review was performed by the project leader, project programmer, and interface designer to identify areas of the preliminary design that had worked successfully and areas that had been problematic. In general, reviewers had little difficulty navigating through the INFeRS menus once they had completed a few searches. Some reviewers were not sure how to start the process of retrieval. In these cases, the observing project member described the following three-step procedure: (1) Define the subset of the datafile to be extracted by selecting the variables of interest and specifying particular values of those variables if desired, (2) Select a delimiting format in which to output the results, and (3) Initiate the search and extraction. Once the general process was outlined, those reviewers initially uncertain about where to begin proceeded to use INFeRS with little hindrance. The most confusing aspects of the interface were the labels of certain menu items and the wording of various prompts and error messages. Reviewers had some difficulty using the interface to the subject index, which was not unexpected. The subject index interface had been the most difficult to design. For that reason, we anticipated there would be some usability problems with the preliminary design.

In accordance with the results of the system review, the interface designer and project programmer modified screen designs to address the difficulties the reviewers had experienced. Changes in wording of confusing menu labels, prompts, and error messages were easily implemented. With regard to the subject index, the nature of the difficulties experienced by reviewers indicated that a significant redesign was needed. However, constraints on time remaining for development prohibited the team from engaging in major redesign efforts. As a compromise, modifications that did not require significant reprogramming of the subject index were made, with the realization that this was not an ideal solution.

Recommendations for interface design

Do not implement database design until user interface requirements have been identified .

The user interface component of an interactive computer system is most likely to change during development. As development proceeds and end-users review the system, new requirements will be identified and problems will be encountered for which the interface design will need modification. In an information retrieval system, the user interface is often tied closely to the structure of the underlying database. Therefore, implementing a database design too early in the development cycle can create problems when interface requirements are not well known. It can be difficult or impossible to make necessary changes to the interface to accommodate user requirements identified later in development. It can also be costly in terms of programming time and effort to change an existing database design once it has been implemented. The best solution is to design the database structure so that access to the data is as independent from the storage of the data as possible. This approach results in a relatively stable database design that can accommodate the dynamic nature of user interface design. The database design can be prototyped or implemented before user interface design is complete while minimizing the risk that it will impose restrictions on later modifications of the interface.

Identify user interface requirements by prototyping interface designs.

Building a prototype, or model, of the user interface of an interactive computer system has a number of benefits. Functional and information requirements of the interface can be identified early in the development cycle and can be factored into the database and system design from the beginning. It is often easier for end-users to evaluate a tangible model of a system against their needs for that system rather than to describe and express those needs intellectually. The old adage "A picture is worth a thousand words" applies here. Another benefit is that technical requirements and constraints can be identified before development of the production system begins. Development time would not be wasted on ideas that cannot be implemented technically. The cost of interface prototyping is that it requires time out of the development schedule. Prototyping is often viewed as a waste of development time because it is essentially a period of experimentation with no guarantee of a usable product. Effective interface prototyping can result in a more usable system. Given the importance of user-friendliness on the effectiveness and success of any computer application, the benefits of prototyping the interface can outweigh its costs.

Separate user interface development and database development.

Effective interface design is dynamic and iterative in nature. Initial designs are prototyped or implemented, evaluated, redesigned, reimplemented, re-evaluated, etc., until a satisfactory design is developed. Implementation of user interfaces typically requires more redesign and reprogramming than any other component of an interactive computer system. Database design and implementation also requires significant development time but is subject to fewer modifications. To accommodate the variable nature of interface and database design and the considerable effort required for their production, user interface and database development should be physically or temporally separated.

Development of the two components can be physically separated by dividing the responsibility for implementation between two or more programmers. The multiple programmer approach permits the user interface and database can be designed and developed concurrently. Frequent modification of the user interface can be done without affecting other system development. Multiple programmers on a single application requires extensive communication and coordination of effort, which can increase management costs over a single programmer approach. A potential benefit of the increased communication required is that the overall system design tends to be more organized.

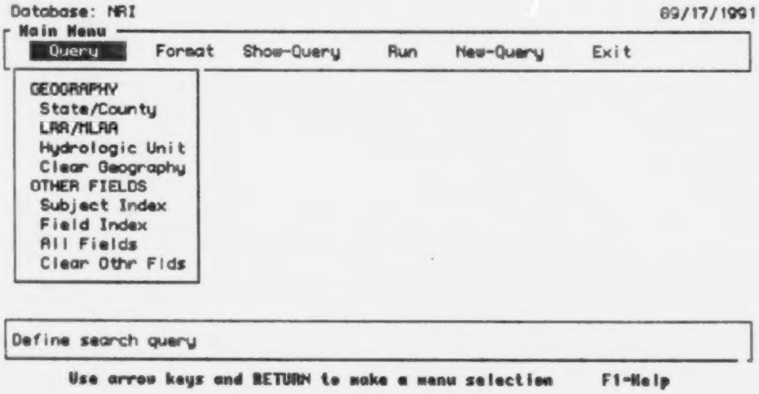
Temporally dividing user interface and database implementation between two phases of system development allows project resources to be concentrated on a single component of the system at one time. User interface and database development each get the attention they require. Management of project resources can be easier as well. In taking this approach, interface development should be done before database development. Given the relative difficulty of modifying a database design once implemented and the dependence of the user interface on database structure, it is important to ensure that functional and information interface requirements are identified at the beginning of the development process.

System Description

INFeRS is a windowed menu system running under VT100 terminal emulation. Interactions are controlled by the arrow keys, the enter (or return) key, the escape key, and the function keys 1-4. The basic interactions are highlighting a choice from the menu and pressing the enter key to send that selection to the system. In certain cases (numeric limiting and word searches) typing is required.

Transactions

INFeRS consists of four major transactions: definition of the search, definition of the format and destination for the output set, running the search and transferring the resulting subset to the local system, and exiting INFeRS. INFeRS also has two optional functions: reviewing the search definitions, and erasing all definitions to begin a new search.



Database: NRI 09/17/1991

Define search query

Use arrow keys and RETURN to make a menu selection F1-Help

figure 1 INFeRS Main Menu

Detailed Description

Search definition

To define a search the user sets limits on the data to be retrieved (geographical, certain values in fields, numeric conditions), and specifies which fields in the records retrieved through the limiting process are to be included in the output file.

The definitions/limits are all determined by the field names, and values within fields. All of the information needed to make those selections is available through the Query window. The information is organized into a series of indexes. In the case of the NRI database (the example used throughout this Section) the indexes are sorted into Geographical and Other Fields.

The minimum definition is a geographic limit and a file format. By default INFeRS would limit the search to that geography and output the default fields (the basic record identifers of sample point information and geographic classifications) to the file format specified by the user.

INDEXES

Geographic index: There are three geographic schemes used in the NRI data: State/County, Land Resource Region/Major Land Resource Area, and Hydrologic Unit. They have been applied as mutually exclusive definitions; the user may only select one scheme at a time.

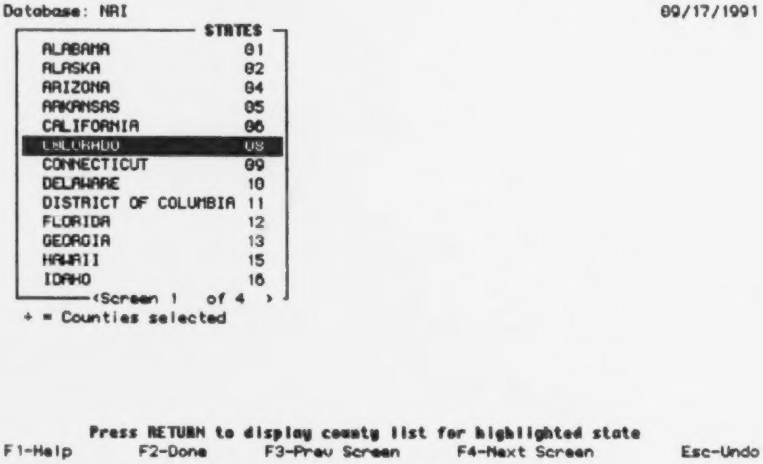
Different interfaces have been used for each scheme, their design was determined by how familiar the user might be with scheme and its underlying concept.

From our interviews with prospective users we concluded that they would always need to set a geographic limit on their search. The NRI database contains over 840,000 records. A patron attempting to dowload the entire database somehow, or limit on some variable value against the entire database, would quit in frustration long before the system could execute a request of that size. Therefore the geographic indexes are listed on the main Query menu rather than imbedded in the field index.

We also decided that certain elements of the record would automatically be output in order to ensure that the user would always be able to identify the records retrieved. The geographic identifications are included in that automatic output so the user does not have to specify that a particular geographic field (state/county, hydrologic unit, or Land Resource Region/Major Land Resource Area) is to be included in the output file.

STATES AND COUNTIES The political concept of states and counties is familiar to most users, as are the actual names of the states and counties. Thus the menu windows are simple lists.

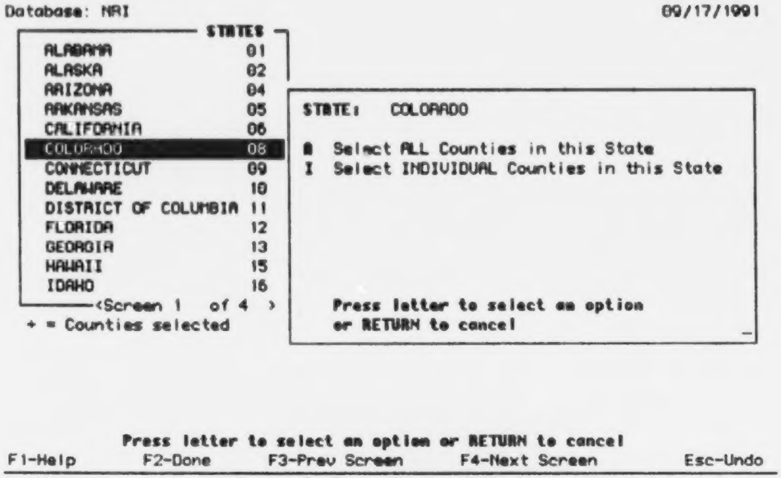
After selecting the State/County option from the main Query menu the user sees a list of states:



```
Database: NRI                                                09/17/1991
             STATES
  ALABAMA                   01
  ALASKA                    02
  ARIZONA                   04
  ARKANSAS                  05
  CALIFORNIA                06
  COLORADO                  08
  CONNECTICUT               09
  DELAWARE                  10
  DISTRICT OF COLUMBIA      11
  FLORIDA                   12
  GEORGIA                   13
  HAWAII                    15
  IDAHO                     16
        <Screen 1   of 4  >
 + = Counties selected



        Press RETURN to display county list for highlighted state
F1-Help      F2-Done      F3-Prev Screen      F4-Next Screen        Esc-Undo
```

figure 2 State List Window

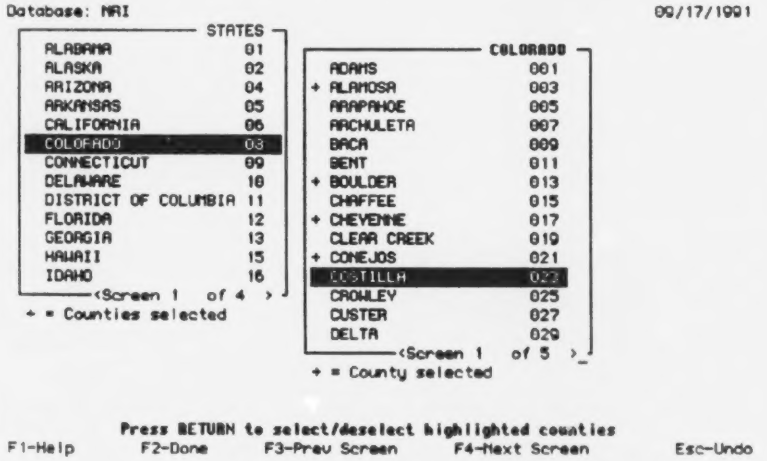
Pressing enter when the highlight is on the state of interest opens an option window:



```
Database: NRI                                              09/17/1991
 ---------------------- STATES -
  ALABAMA                  01
  ALASKA                   02
  ARIZONA                  04
  ARKANSAS                 05   STATE:   COLORADO
  CALIFORNIA               06
  COLORADO                 08   A  Select ALL Counties in this State
  CONNECTICUT              09   I  Select INDIVIDUAL Counties in this State
  DELAWARE                 10
  DISTRICT OF COLUMBIA     11
  FLORIDA                  12
  GEORGIA                  13
  HAWAII                   15
  IDAHO                    16
 -----------<Screen 1   of 4  >     Press letter to select an option
 + = Counties selected              or RETURN to cancel

             Press letter to select an option or RETURN to cancel
F1-Help      F2-Done      F3-Prev Screen      F4-Next Screen       Esc-Undo
```

figure 3 State Option Window

The user can limit the search to all counties in the state, or select individual counties. If they type I for individual counties they see a list of the counties in their state. Again they highlight and press Enter to mark the counties for which they want data. Each county selected is flagged with an asterisk. Highlighting a marked county and pressing Enter deletes that county from the selected list.



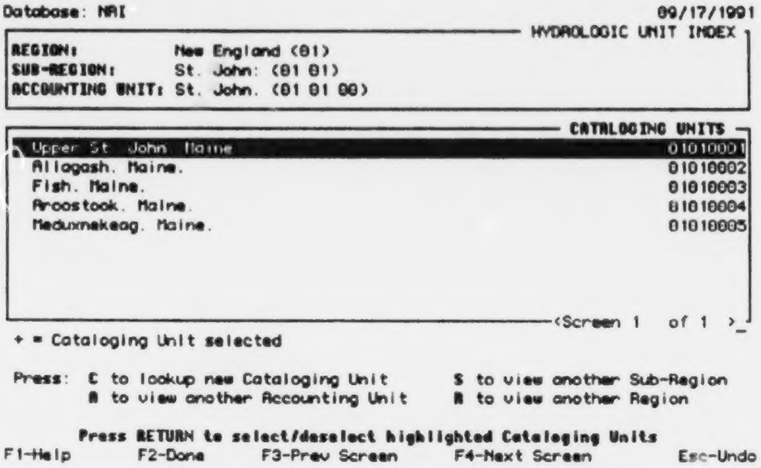
```
Database: NRI                                              09/17/1991
 ---------------------- STATES -
  ALABAMA                  01    -------------------- COLORADO -
  ALASKA                   02      ADAMS                  001
  ARIZONA                  04    + ALAMOSA                003
  ARKANSAS                 05      ARAPAHOE               005
  CALIFORNIA               06      ARCHULETA              007
  COLORADO                 08      BACA                   009
  CONNECTICUT              09      BENT                   011
  DELAWARE                 10    + BOULDER                013
  DISTRICT OF COLUMBIA     11      CHAFFEE                015
  FLORIDA                  12    + CHEYENNE               017
  GEORGIA                  13      CLEAR CREEK            019
  HAWAII                   15    + CONEJOS                021
  IDAHO                    16      COSTILLA               023
 -----------<Screen 1   of 4  >    CROWLEY                025
 + = Counties selected             CUSTER                 027
                                   DELTA                  029
                                 -------------<Screen 1   of 5  >
                                 + = County selected

             Press RETURN to select/deselect highlighted counties
F1-Help      F2-Done      F3-Prev Screen      F4-Next Screen       Esc-Undo
```

figure 4 County Selection Window

HYDROLOGIC UNITS In the case of the hydrologic units we assumed a user wanting to limit their search to a hydrological region would be familiar with the concept of a watershed, but might not know the particular name and/or hierarchical level the U.S. Geological Survey (the creator of the scheme) had assigned to the desired region.

The Survey divides the country into Regions, subdivides the Regions into Sub-Regions, further divides the Sub-Regions into Accounting Units, and finally divides Accounting Units into Cataloging Units, watersheds of over 700 square miles. Thus the menu system offers a word lookup and browse of the four-level hierarchy.

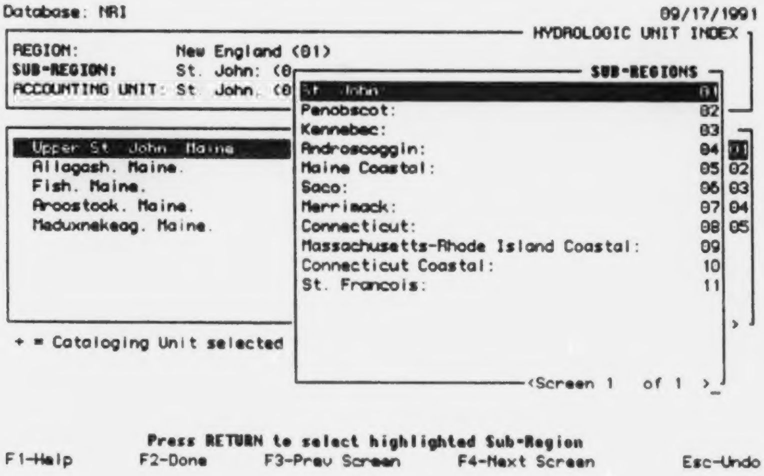
When the user selects Hydrologic Unit for geography they see the Cataloging Units for the first Accounting Unit/Sub-Region/Region in the hierarchy



```
Database: NRI                                                  09/17/1991
┌──────────────────────────────────────────── HYDROLOGIC UNIT INDEX ┐
│REGION:            New England (01)                                │
│SUB-REGION:        St. John: (01 01)                               │
│ACCOUNTING UNIT:   St. John. (01 01 00)                            │
└───────────────────────────────────────────────────────────────────┘
┌──────────────────────────────────────────────── CATALOGING UNITS ─┐
│ Upper St  John  Maine                                     01010001│
│ Allagash. Maine.                                          01010002│
│ Fish. Maine.                                              01010003│
│ Aroostook. Maine.                                         01010004│
│ Meduxnekeag. Maine.                                       01010005│
│                                                                   │
│                                                                   │
└─────────────────────────────────────────────<Screen 1   of 1  >_┘
 + = Cataloging Unit selected

 Press:  C to lookup new Cataloging Unit        S to view another Sub-Region
         A to view another Accounting Unit      R to view another Region

          Press RETURN to select/deselect highlighted Cataloging Units
F1-Help      F2-Done      F3-Prev Screen      F4-Next Screen         Esc-Undo
```

figure 5 First Screen Hydrologic Geography

They have the option to highlight the Cataloging Unit from this screen, or they can travel through the hierarchy. One letter commands display other Accounting Units in the same Sub-Region, and other Sub-Regions in the same Region,

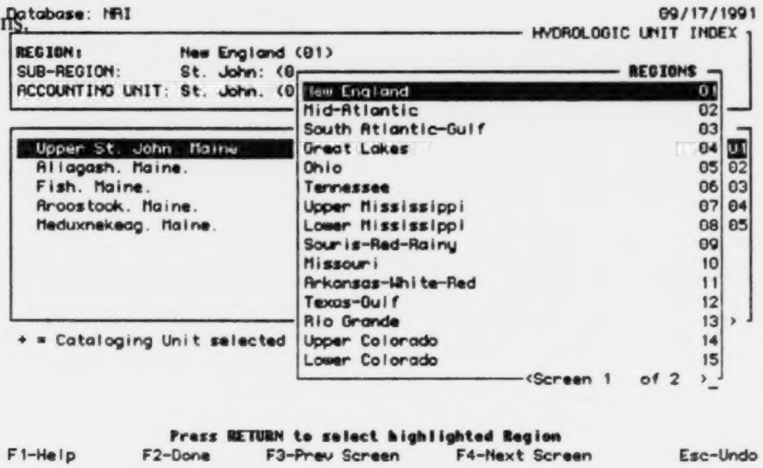


```
Database: NRI                                                  09/17/1991
┌──────────────────────────────────────────── HYDROLOGIC UNIT INDEX ┐
│REGION:            New England (01)                                │
│SUB-REGION:        St. John: (0┌──────────────────── SUB-REGIONS ─┐│
│ACCOUNTING UNIT:   St. John. (0│St  John                         01│┘
└─────────────────────────────── │Penobscot:                       02│─
┌─────────────────────────────── │Kennebec:                        03│ ┐
│ Upper St  John  Maine          │Androscoggin:                    04│01
│ Allagash. Maine.               │Maine Coastal:                   05│02
│ Fish. Maine.                   │Saco:                            06│03
│ Aroostook. Maine.              │Merrimack:                       07│04
│ Meduxnekeag. Maine.            │Connecticut:                     08│05
│                                │Massachusetts-Rhode Island Coastal: 09│
│                                │Connecticut Coastal:             10│
│                                │St. Francois:                    11│
└─────────────────────────────── │                                   │ > ┘
 + = Cataloging Unit selected    │                                   │
                                 └──────────────<Screen 1   of 1  >_┘

               Press RETURN to select highlighted Sub-Region
F1-Help      F2-Done      F3-Prev Screen      F4-Next Screen         Esc-Undo
```

figure 6 Sub-Region List

BEST COPY AVAILABLE

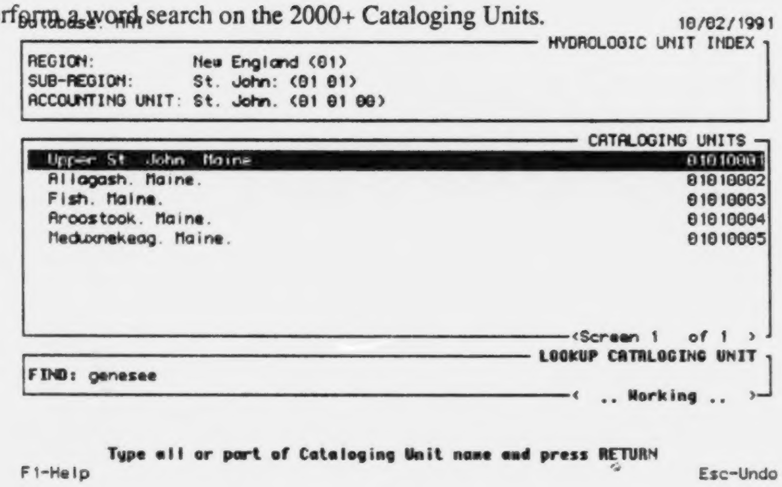
or the list of Regions.



```
Database: NRI                                                    09/17/1991
┌──────────────────────────────────────────────── HYDROLOGIC UNIT INDEX ┐
│REGION:            New England (01)                                    │
│SUB-REGION:        St. John: (0┌────────────────────────── REGIONS ┐   │
│ACCOUNTING UNIT: St. John. (0 │New England                        01│───┘
└──────────────────────────────│Mid-Atlantic                       02│
┌──────────────────────────────│South Atlantic-Gulf                03│ ┐
│ Upper St. John  Maine        │Great Lakes                        04│01
│ Allagash. Maine.             │Ohio                               05│02
│ Fish. Maine.                 │Tennessee                          06│03
│ Aroostook. Maine.            │Upper Mississippi                  07│04
│ Meduxnekeag. Maine.          │Lower Mississippi                  08│05
│                              │Souris-Red-Rainy                   09│
│                              │Missouri                           10│
│                              │Arkansas-White-Red                 11│
│                              │Texas-Gulf                         12│
└──────────────────────────────│Rio Grande                         13│ > ┘
 + = Cataloging Unit selected  │Upper Colorado                     14│
                               │Lower Colorado                     15│
                               └──────────────<Screen 1   of 2  >_┘

                  Press RETURN to select highlighted Region
F1-Help      F2-Done      F3-Prev Screen      F4-Next Screen       Esc-Undo
```

figure 7 Region List

They can also perform a word search on the 2000+ Cataloging Units.



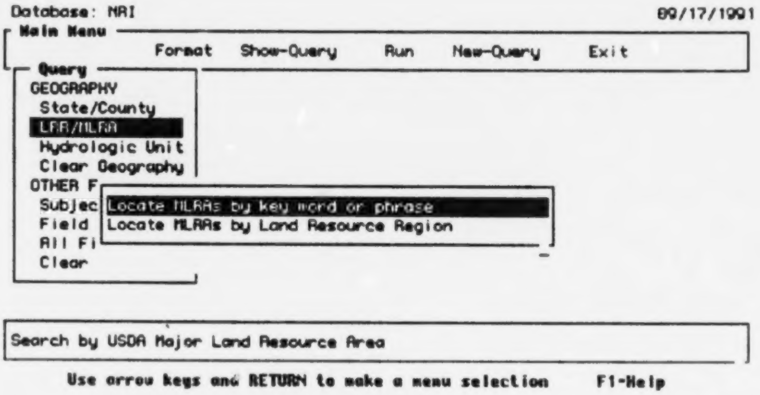
```
Database: NRI                                                    10/02/1991
┌──────────────────────────────────────────────── HYDROLOGIC UNIT INDEX ┐
│REGION:            New England (01)                                    │
│SUB-REGION:        St. John: (01 01)                                   │
│ACCOUNTING UNIT: St. John. (01 01 00)                                  │
└───────────────────────────────────────────────────────────────────────┘
┌──────────────────────────────────────────────────── CATALOGING UNITS ┐
│ Upper St  John  Maine                                         01010001│
│ Allagash. Maine.                                              01010002│
│ Fish. Maine.                                                  01010003│
│ Aroostook. Maine.                                             01010004│
│ Meduxnekeag. Maine.                                           01010005│
│                                                                       │
│                                                                       │
└─────────────────────────────────────────────────<Screen 1   of 1  > ┘
┌─────────────────────────────────────────────── LOOKUP CATALOGING UNIT ┐
│FIND: genesee                                                          │
└─────────────────────────────────────────────────<  .. Working ..  >─┘

           Type all or part of Cataloging Unit name and press RETURN
F1-Help                                                             Esc-Undo
```

figure 8 Cataloging Unit word lookup

MAJOR LAND RESOURCE AREAS The U.S. Department of Agriculture uses another hierarchical classification scheme, based on the land "as a resource for farming, ranching, forestry, engineering, recreation, and other uses." (Land Resource, 1981) They divide the country into land resource units "characterized by a particular pattern of soils, climate, water resources, and land uses." (Land Resource, 1981)[1] Those regions are grouped into Major Land Resource Areas, and further clustered into Land Resource Regions. According to our focus group, only researchers using other USDA data are familiar with, and would need, this classification. Therefore INFeRS does not initially display any list, instead users are given the choice of doing a word search for a MLRA or looking at a list.

---

[1] Land Resource Regions and Major Land Resource Areas of The United States. *Agriculture handbook* 296 Washington, D.C.: U.S. Department of Agriculture, Soil Conservation Service, 1981



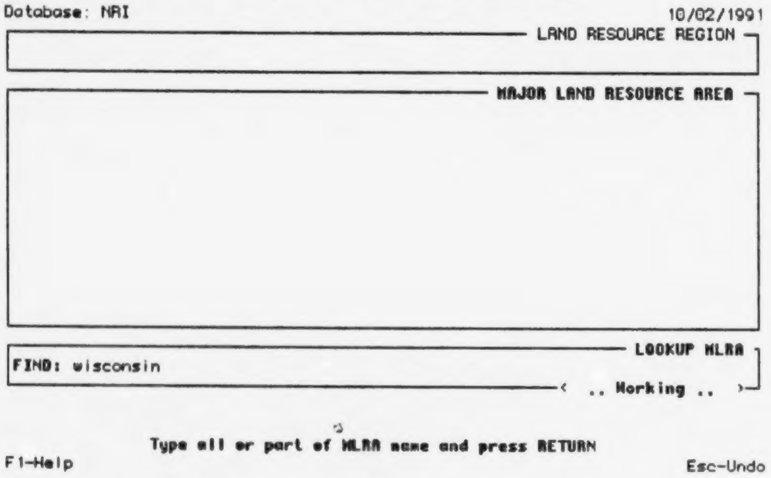
Database: NRI 09/17/1991

Search by USDA Major Land Resource Area

Use arrow keys and RETURN to make a menu selection F1-Help

figure 9 LRR/MLRA Option Window

Users who are familiar with the classification would presumably want to search by word. They are presented with a lookup-window where they type in a word from their Area.



Database: NRI 10/02/1991

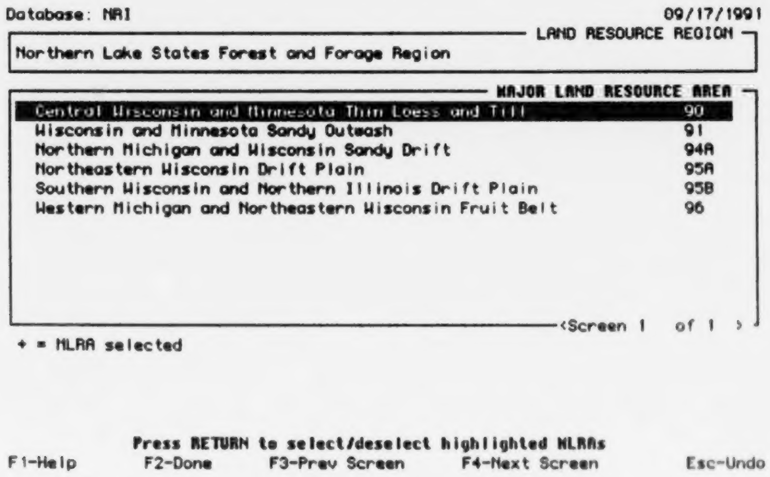
FIND: wisconsin

.. Working ..

Type all or part of MLRA name and press RETURN

figure 10 LRR/MLRA Word Look-Up Screen

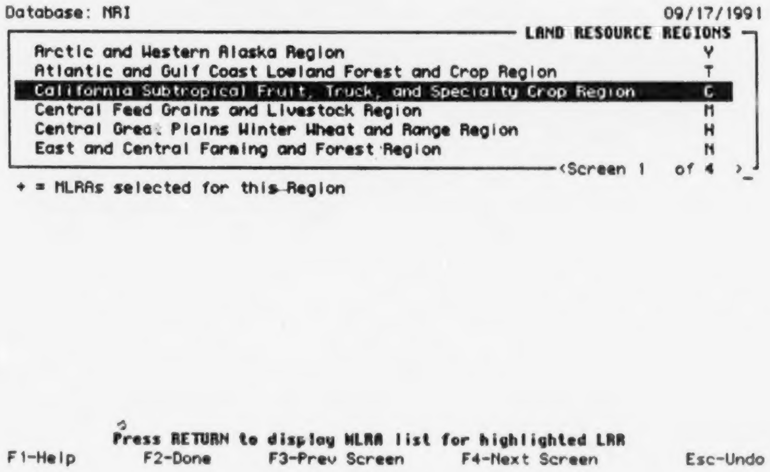
The system displays the MLRAs that meet the word search. The Land Resource Region for the particular MLRA highlighted displays in the upper window. They then highlight and press Enter for any MLRAs of interest.



```
Database: NRI                                                09/17/1991
                                             LAND RESOURCE REGION
Northern Lake States Forest and Forage Region

                                         MAJOR LAND RESOURCE AREA
 Central Wisconsin and Minnesota Thin Loess and Till              90
 Wisconsin and Minnesota Sandy Outwash                            91
 Northern Michigan and Wisconsin Sandy Drift                      94A
 Northeastern Wisconsin Drift Plain                               95A
 Southern Wisconsin and Northern Illinois Drift Plain             95B
 Western Michigan and Northeastern Wisconsin Fruit Belt           96




                                                <Screen 1    of 1  >
 + = MLRA selected



              Press RETURN to select/deselect highlighted MLRAs
F1-Help       F2-Done      F3-Prev Screen     F4-Next Screen        Esc-Undo
```

figure 11 MLRA Word Search Matches

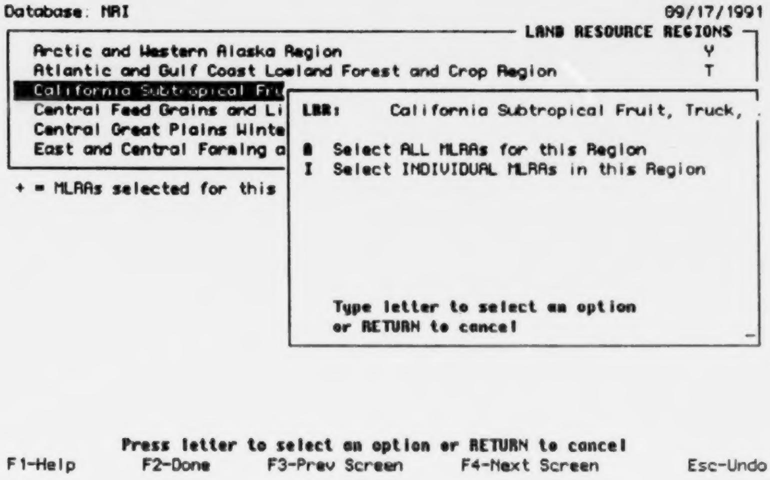
If they wish to browse the complete list they select the second option: Locate MLRAs by Land Resource Region. Here the interface parallels the State/County selection sequence. INFeRS displays the LRRs in alphabetical order.



```
Database: NRI                                                09/17/1991
                                            LAND RESOURCE REGIONS
 Arctic and Western Alaska Region                                 Y
 Atlantic and Gulf Coast Lowland Forest and Crop Region           T
 California Subtropical Fruit, Truck, and Specialty Crop Region   C
 Central Feed Grains and Livestock Region                         M
 Central Great Plains Winter Wheat and Range Region               H
 East and Central Farming and Forest Region                       N
                                                <Screen 1    of 4  >
 + = MLRAs selected for this Region









             Press RETURN to display MLRA list for highlighted LRR
F1-Help       F2-Done      F3-Prev Screen     F4-Next Screen        Esc-Undo
```

figure 12 Land Resource Regions Window

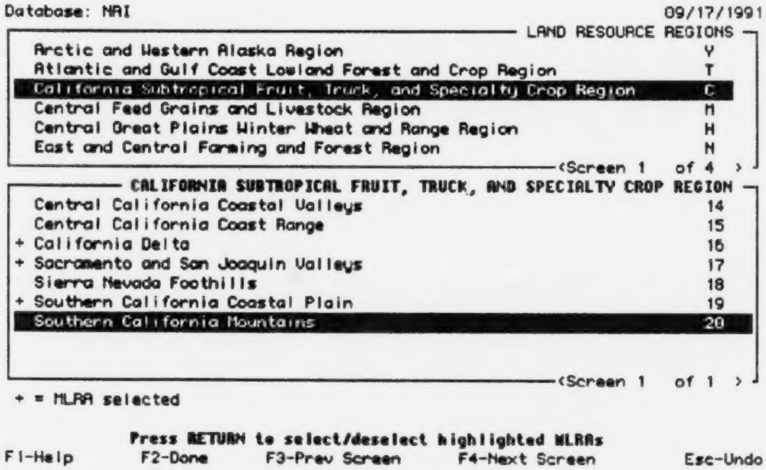
The user selects a LRR and is given the choice of selecting all the MLRAs in that Region, or selecting from a list of the individual MLRAs.



```
Database: NRI                                                    09/17/1991
┌──────────────────────────────────────────────── LAND RESOURCE REGIONS ─┐
│ Arctic and Western Alaska Region                                      Y │
│ Atlantic and Gulf Coast Lowland Forest and Crop Region                T │
│ California Subtropical Fru┌─────────────────────────────────────────────┐
│ Central Feed Grains and Li│ LRR:     California Subtropical Fruit, Truck, .
│ Central Great Plains Winte│                                             │
│ East and Central Farming a│ A  Select ALL MLRAs for this Region         │
└───────────────────────────│ I  Select INDIVIDUAL MLRAs in this Region   │
 + = MLRAs selected for this│                                             │
                            │                                             │
                            │    Type letter to select an option          │
                            │    or RETURN to cancel                     _│
                            └─────────────────────────────────────────────┘

           Press letter to select an option or RETURN to cancel
F1-Help     F2-Done     F3-Prev Screen     F4-Next Screen          Esc-Undo
```

figure 13 Land Resource Regions Option Window

Selecting individual MLRAs works like other selection windows, the Enter key selects, and de-selects the highlighted value. Selected values are marked with an asterisk.



```
Database: NRI                                                    09/17/1991
┌──────────────────────────────────────────────── LAND RESOURCE REGIONS ─┐
│ Arctic and Western Alaska Region                                      Y │
│ Atlantic and Gulf Coast Lowland Forest and Crop Region                T │
│ California Subtropical Fruit, Truck, and Specialty Crop Region        C │
│ Central Feed Grains and Livestock Region                              M │
│ Central Great Plains Winter Wheat and Range Region                     H │
│ East and Central Farming and Forest Region                            N │
└──────────────────────────────────────────────────<Screen 1   of 4   >─┘
┌────── CALIFORNIA SUBTROPICAL FRUIT, TRUCK, AND SPECIALTY CROP REGION ─┐
│ Central California Coastal Valleys                                   14 │
│ Central California Coast Range                                       15 │
│+ California Delta                                                    16 │
│+ Sacramento and San Joaquin Valleys                                  17 │
│ Sierra Nevada Foothills                                              18 │
│+ Southern California Coastal Plain                                   19 │
│ Southern California Mountains                                        20 │
│                                                                         │
└──────────────────────────────────────────────────<Screen 1   of 1   >─┘
 + = MLRA selected

           Press RETURN to select/deselect highlighted MLRAs
F1-Help     F2-Done     F3-Prev Screen     F4-Next Screen          Esc-Undo
```

figure 14 Major Land Resource Area Window

Other indexes Search conditions on non-geography fields are all optional. Using the indexes the user can specify that only records with a certain value in one or more fields be selected, and they can specify which fields should be included in the output file.

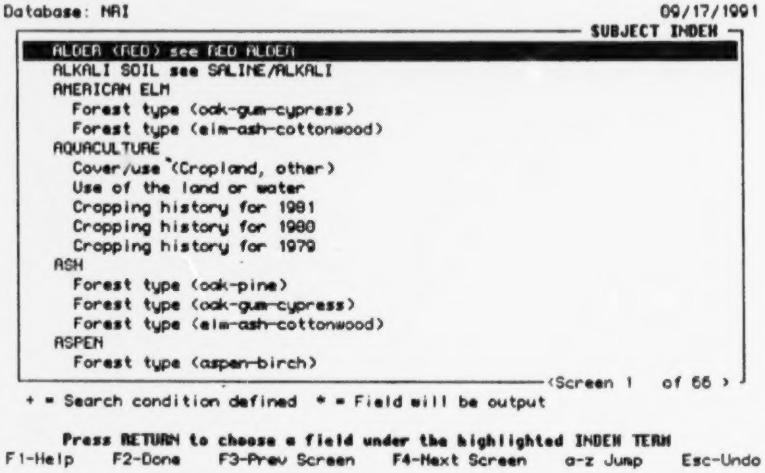
Both indexes cover the non-geographic fields. In the case of the NRI database almost 70 fields are indexed. Users can approach the list in two ways: the Subject Index and the Field Index.

Finding aids to a dataset should be designed for users unfamiliar with the contents of a file, and those who have used the file before (either in print or computer forms). Users unfamiliar with the contents of the database need a detailed index listing all the significant concepts in the file, regardless of their location (as a field name, or as a definition of a code used in the field). The Subject Index was created to fill that need.

However, users familiar with the file, the fields, and the possible codes in the fields, would not want to navigate through the extensive Subject Index to define their search. The Field Index was created for their purposes.

Essentially the two indexes parallel the finding aids of a book. The Field Index is a table of contents, and the Subject Index is the back index. But, unlike print, INFeRS looks up the topics for the user.

SUBJECT INDEX The Subject Index is an alphabetical keyword list of the fields and the value descriptions within fields.



```
Database: NRI                                                09/17/1991
                                                          SUBJECT INDEX
    ALDER (RED) see RED ALDER
    ALKALI SOIL see SALINE/ALKALI
    AMERICAN ELM
      Forest type (oak-gum-cypress)
      Forest type (elm-ash-cottonwood)
    AQUACULTURE
      Cover/use (Cropland, other)
      Use of the land or water
      Cropping history for 1981
      Cropping history for 1980
      Cropping history for 1979
    ASH
      Forest type (oak-pine)
      Forest type (oak-gum-cypress)
      Forest type (elm-ash-cottonwood)
    ASPEN
      Forest type (aspen-birch)
                                                    <Screen 1   of 66 >
  + = Search condition defined  * = Field will be output

       Press RETURN to choose a field under the highlighted INDEX TERM
F1-Help   F2-Done   F3-Prev Screen   F4-Next Screen   a-z Jump   Esc-Undo
```

figure 15 First Screen — Subject Index

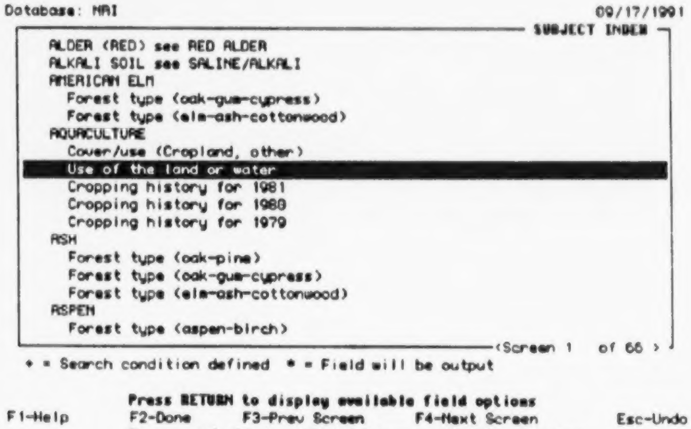
For example field 21 in the NRI database is Cropping History. Two pages of coded values in the documentation describe the possible values for field 21.

Code AE139 stands for Cropland, Other — Aquaculture.

Code AE120 stands for Cropland, Other — Summer fallow.

Code GA630 stands for Permanent snow and ice fields, etc.

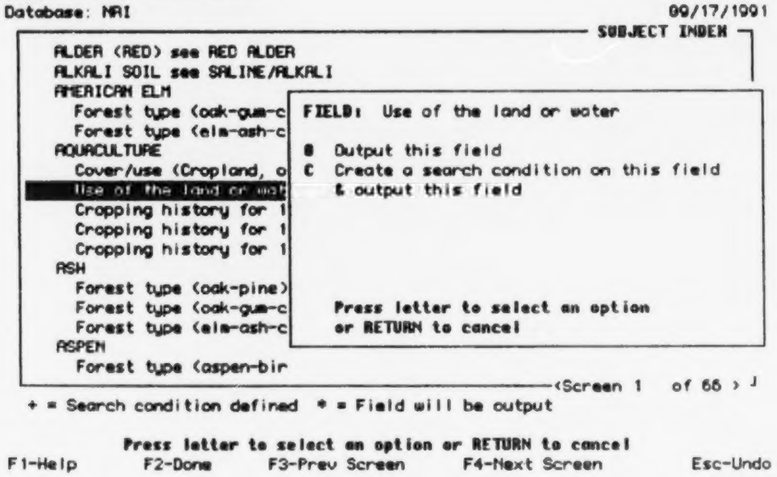
The subject index includes the heading Aquaculture, with entries for the specific value/field where aquaculture appears. The user selects the heading Aquaculture by pressing Enter at the highlight. The highlight then jumps into the value/field list within the subject heading.



```
Database: NRI                                                09/17/1991
                                                          SUBJECT INDEX
    ALDER (RED) see RED ALDER
    ALKALI SOIL see SALINE/ALKALI
    AMERICAN ELM
      Forest type (oak-gum-cypress)
      Forest type (elm-ash-cottonwood)
    AQUACULTURE
      Cover/use (Cropland, other)
      Use of the land or water
      Cropping history for 1981
      Cropping history for 1980
      Cropping history for 1979
    ASH
      Forest type (oak-pine)
      Forest type (oak-gum-cypress)
      Forest type (elm-ash-cottonwood)
    ASPEN
      Forest type (aspen-birch)
                                                    <Screen 1   of 66 >
  + = Search condition defined  * = Field will be output

             Press RETURN to display available field options
F1-Help     F2-Done     F3-Prev Screen     F4-Next Screen          Esc-Undo
```

figure 16 Value/Field Line in Subject Index

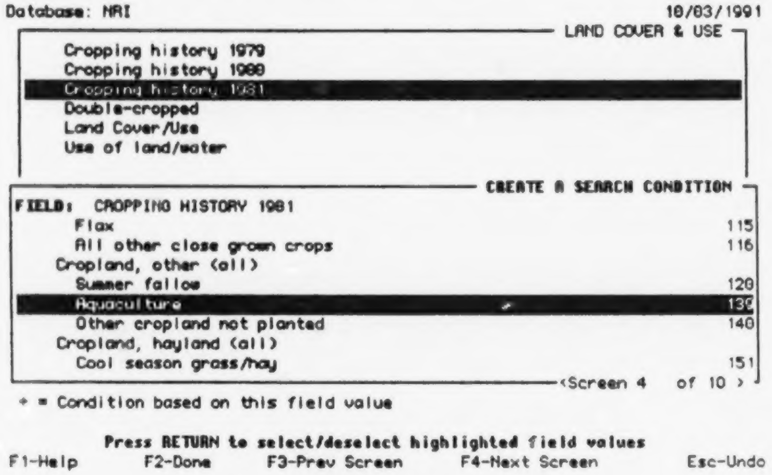
Another Enter displays the option window.



```
Database: NRI                                                  09/17/1991
┌─────────────────────────────────────────────────── SUBJECT INDEX ┐
   ALDER (RED) see RED ALDER                                         │
   ALKALI SOIL see SALINE/ALKALI                                     │
   AMERICAN ELM               ┌──────────────────────────────────────┐
     Forest type (oak-gum-c   │FIELD: Use of the land or water       │
     Forest type (elm-ash-c   │                                      │
   AQUACULTURE                │O  Output this field                  │
     Cover/use (Cropland, o   │C  Create a search condition on this field
     Use of the land or wat   │   & output this field                │
     Cropping history for 1   │                                      │
     Cropping history for 1   │                                      │
     Cropping history for 1   │                                      │
   ASH                        │                                      │
     Forest type (oak-pine)   │                                      │
     Forest type (oak-gum-c   │   Press letter to select an option   │
     Forest type (elm-ash-c   │   or RETURN to cancel                │
   ASPEN                      └──────────────────────────────────────┘
     Forest type (aspen-bir
└──────────────────────────────────────────<Screen 1   of 66 > ┘
 + = Search condition defined  * = Field will be output

          Press letter to select an option or RETURN to cancel
F1-Help      F2-Done      F3-Prev Screen      F4-Next Screen      Esc-Undo
```

figure 17 Field Option Window

The user can select this field to be included in the output file. Or they can specify search conditions on the field, which automatically flags it for output. If the user selects the former option all occurrences of the field are marked with an asterisk. If the user selects the latter INFeRS displays the list of value descriptions with the value related to the subject heading highlighted.

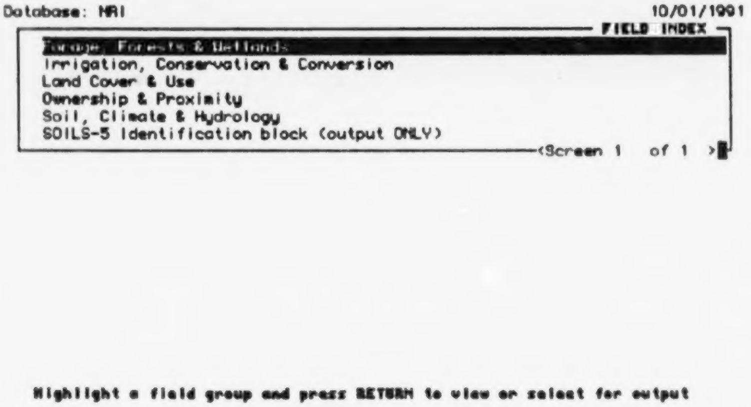


```
Database: NRI                                                  10/03/1991
┌─────────────────────────────────────────────── LAND COVER & USE ┐
     Cropping history 1979                                          │
     Cropping history 1980                                          │
     Cropping history 1981                                          │
     Double-cropped                                                 │
     Land Cover/Use                                                 │
     Use of land/water                                              │
│                                                                   │
┌─────────────────────────────────────────── CREATE A SEARCH CONDITION ┐
FIELD:  CROPPING HISTORY 1981
      Flax                                                        115│
      All other close grown crops                                 116│
    Cropland, other (all)
      Summer fallow                                               120│
      Aquaculture                                                 130│
      Other cropland not planted                                  140│
    Cropland, hayland (all)
      Cool season grass/hay                                       151│
└──────────────────────────────────────────<Screen 4   of 10 > ┘
 + = Condition based on this field value

         Press RETURN to select/deselect highlighted field values
F1-Help      F2-Done      F3-Prev Screen      F4-Next Screen      Esc-Undo
```

figure 18 Value/field Selection Window

Users may select as many fields for inclusion in the output file as they want, they may also use the All Fields menu option. Users may also set as many conditions on the search as they want, but these conditions are *cumulative,* they are successive narrowings of the search. Each record must meet all the search conditions. From our interviews with users it is evident that very few searchers will set more than one or two conditions. Thus very few users will limit their search down to zero records.

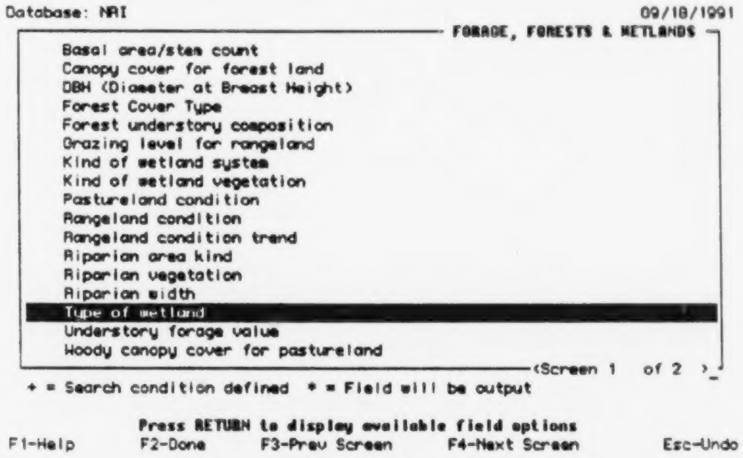
FIELD INDEX In the Field Index each of the fields available for searching has been sorted into a general subject category.



```
Database: NRI                                                10/01/1991
 ──────────────────────────────────────────────────── FIELD INDEX ─┐
│Forage, Forests & Wetlands                                         │
│Irrigation, Conservation & Conversion                              │
│Land Cover & Use                                                   │
│Ownership & Proximity                                              │
│Soil, Climate & Hydrology                                          │
│SOILS-5 Identification block (output ONLY)                         │
└───────────────────────────────────────────<Screen 1   of 1  >┘



        Highlight a field group and press RETURN to view or select for output
```

figure 19 Field Index Category List

Selecting a category displays a list of fields (alphabetically by field name) within the category.

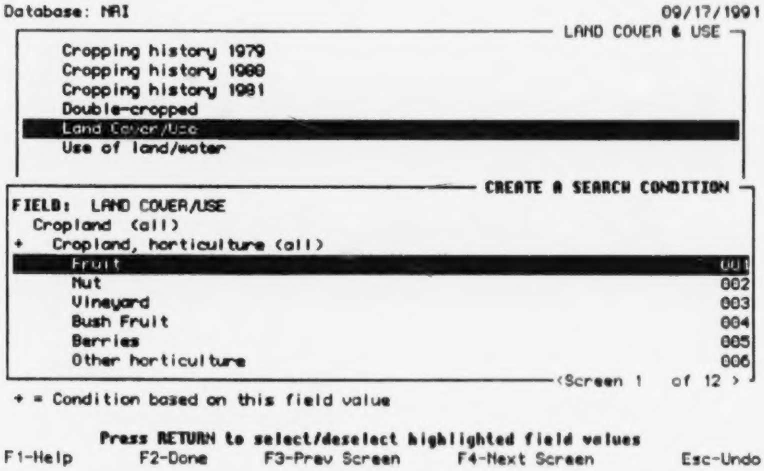


```
Database: NRI                                                09/18/1991
 ─────────────────────────────────── FORAGE, FORESTS & WETLANDS ─┐
│     Basal area/stem count                                       │
│     Canopy cover for forest land                                │
│     DBH (Diameter at Breast Height)                             │
│     Forest Cover Type                                           │
│     Forest understory composition                               │
│     Grazing level for rangeland                                 │
│     Kind of wetland system                                      │
│     Kind of wetland vegetation                                  │
│     Pastureland condition                                       │
│     Rangeland condition                                         │
│     Rangeland condition trend                                   │
│     Riparian area kind                                          │
│     Riparian vegetation                                         │
│     Riparian width                                              │
│     Type of wetland                                             │
│     Understory forage value                                     │
│     Woody canopy cover for pastureland                          │
└────────────────────────────────────────<Screen 1   of 2  >┘
  + = Search condition defined   * = Field will be output

               Press RETURN to display available field options
F1-Help       F2-Done      F3-Prev Screen      F4-Next Screen        Esc-Undo
```

figure 20 List of Fields within a Cluster

Once the user selects a field INFeRS displays the same field selection windows as are displayed from the Subject Index.

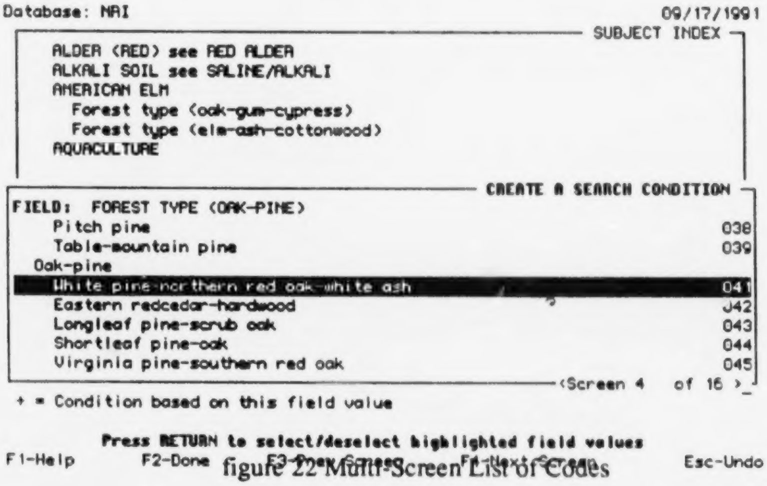
FIELD SELECTION WINDOWS/SEARCH CONDITIONS Again the NRI database provides the best examples of the range of Field Selection Windows developed for INFeRS. A user setting search conditions on a field may select a value or values from a one-screen list.



```
Database: NRI                                                09/17/1991
 ──────────────────────────────────────────────── LAND COVER & USE ─┐
      Cropping history 1979                                          │
      Cropping history 1980                                          │
      Cropping history 1981                                          │
      Double-cropped                                                 │
      Land Cover/Use                                                 │
      Use of land/water                                              │
                                                                     │
 ───────────────────────────────────────── CREATE A SEARCH CONDITION ─┐
FIELD:  LAND COVER/USE
  Cropland  (all)
+   Cropland, horticulture (all)
       Fruit                                                        001
       Nut                                                          002
       Vineyard                                                     003
       Bush Fruit                                                   004
       Berries                                                      005
       Other horticulture                                           006
 ─────────────────────────────────────────────── <Screen 1   of 12 > ┘
 + = Condition based on this field value

          Press RETURN to select/deselect highlighted field values
F1-Help       F2-Done      F3-Prev Screen      F4-Next Screen       Esc-Undo
```

figure 21 One-Screen List of Codes

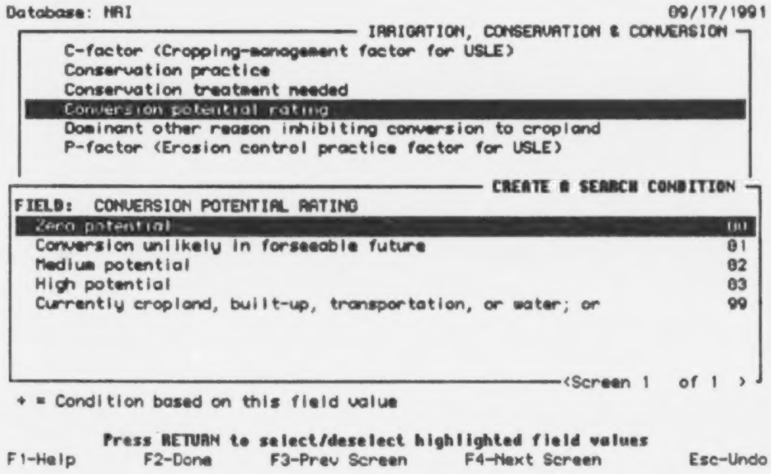
Or they may select one or more values from a multi-screen list.



```
Database: NRI                                                09/17/1991
 ──────────────────────────────────────────────────── SUBJECT INDEX ─┐
    ALDER (RED) see RED ALDER                                        │
    ALKALI SOIL see SALINE/ALKALI                                    │
    AMERICAN ELM                                                     │
      Forest type (oak-gum-cypress)                                  │
      Forest type (elm-ash-cottonwood)                               │
    AQUACULTURE                                                      │
                                                                     │
 ───────────────────────────────────────── CREATE A SEARCH CONDITION ─┐
FIELD:  FOREST TYPE (OAK-PINE)
    Pitch pine                                                      038
    Table-mountain pine                                             039
  Oak-pine
    White pine-northern red oak-white ash                           041
    Eastern redcedar-hardwood                                       042
    Longleaf pine-scrub oak                                         043
    Shortleaf pine-oak                                              044
    Virginia pine-southern red oak                                  045
 ─────────────────────────────────────────────── <Screen 4   of 16 >_┘
 + = Condition based on this field value

          Press RETURN to select/deselect highlighted field values
F1-Help       F2-Done      F3-Prev Screen      F4-Next Screen       Esc-Undo
```

figure 22 Multi-Screen List of Codes

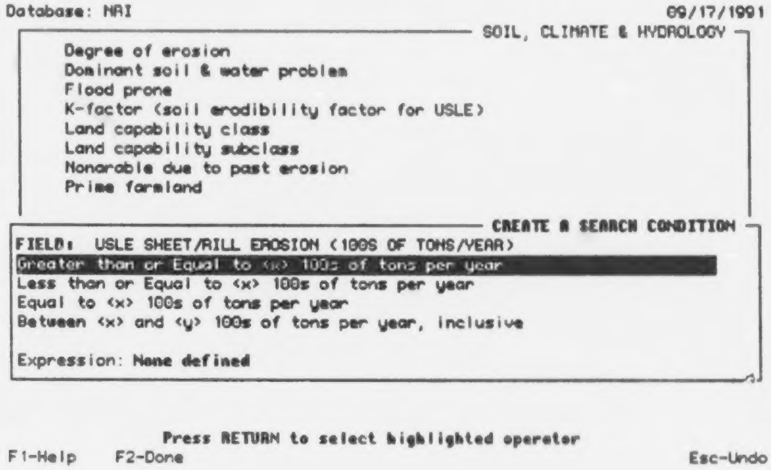
They can also select at different levels, from a hierarchically coded list.



```
Database: NRI                                                    09/17/1991
  ┌────────────────────────── IRRIGATION, CONSERVATION & CONVERSION ─┐
  │   C-factor (Cropping-management factor for USLE)                  │
  │   Conservation practice                                           │
  │   Conservation treatment needed                                   │
  │   Conversion potential rating                                     │
  │   Dominant other reason inhibiting conversion to cropland         │
  │   P-factor (Erosion control practice factor for USLE)             │
 ┌────────────────────────────────────────── CREATE A SEARCH CONDITION ─┐
 │FIELD:  CONVERSION POTENTIAL RATING                                   │
 │ Zero potential                                                    00 │
 │ Conversion unlikely in forseeable future                          01 │
 │ Medium potential                                                  02 │
 │ High potential                                                    03 │
 │ Currently cropland, built-up, transportation, or water; or        99 │
 │                                                                      │
 └─────────────────────────────────────────────<Screen 1   of 1  >──────┘
 + = Condition based on this field value

              Press RETURN to select/deselect highlighted field values
F1-Help        F2-Done        F3-Prev Screen        F4-Next Screen       Esc-Undo
```

figure 23 Hierarchical List of Codes

Finally, they can set numeric limits through a two-screen process. First they select the numeric relationship.

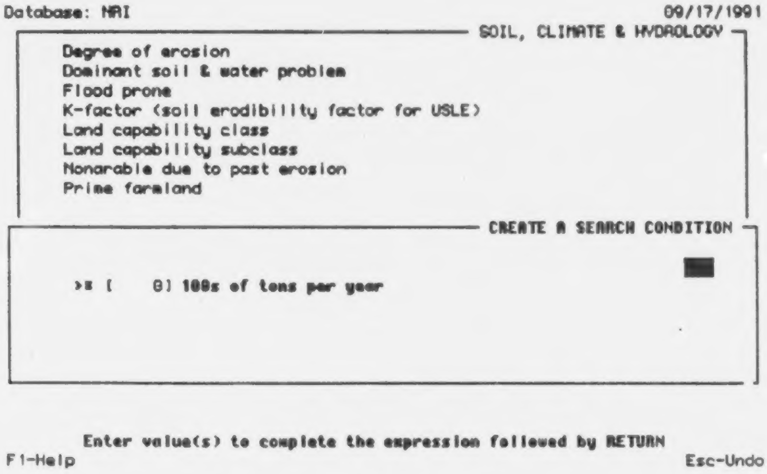


```
Database: NRI                                                    09/17/1991
  ┌──────────────────────────────────────── SOIL, CLIMATE & HYDROLOGY ─┐
  │   Degree of erosion                                                 │
  │   Dominant soil & water problem                                     │
  │   Flood prone                                                       │
  │   K-factor (soil erodibility factor for USLE)                       │
  │   Land capability class                                             │
  │   Land capability subclass                                          │
  │   Nonarable due to past erosion                                     │
  │   Prime farmland                                                    │
 ┌────────────────────────────────────────── CREATE A SEARCH CONDITION ─┐
 │FIELD:  USLE SHEET/RILL EROSION (100S OF TONS/YEAR)                   │
 │Greater than or Equal to <x> 100s of tons per year                    │
 │Less than or Equal to <x> 100s of tons per year                       │
 │Equal to <x> 100s of tons per year                                    │
 │Between <x> and <y> 100s of tons per year, inclusive                  │
 │                                                                      │
 │Expression: None defined                                              │
 └──────────────────────────────────────────────────────────────────────┘

                    Press RETURN to select highlighted operator
F1-Help     F2-Done                                                      Esc-Undo
```

figure 24 Numeric Operators Window

BEST COPY AVAILABLE

Then they fill in the numeric condition. The system displays the units of measure from the codebook.



```
Database: NRI                                                  09/17/1991
 ____________________________________________ SOIL, CLIMATE & HYDROLOGY _
|      Degree of erosion                                                 |
|      Dominant soil & water problem                                     |
|      Flood prone                                                       |
|      K-factor (soil erodibility factor for USLE)                       |
|      Land capability class                                             |
|      Land capability subclass                                          |
|      Nonarable due to past erosion                                     |
|      Prime farmland                                                    |
 ___________________________________________ CREATE A SEARCH CONDITION __
|                                                                         |
|     >= [     0] 100s of tons per year                                   |
|                                                                         |
|_________________________________________________________________________|

        Enter value(s) to complete the expression followed by RETURN
F1-Help                                                          Esc-Undo
```

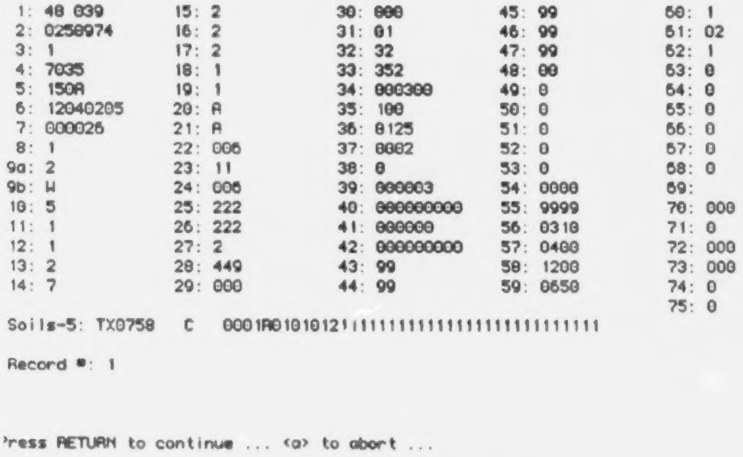
figure 25 Numeric Condition Window

These are only representative examples. There are range-permissible numeric fields, complex fields, and dependent fields which are discussed in the Data Section in the detailed discussion of data coding schemes.

ALL FIELDS As mentioned before, the default output file is only a subset of the whole record. If a user wants all the fields output they select the All Fields Option from the Main Menu.

Format options

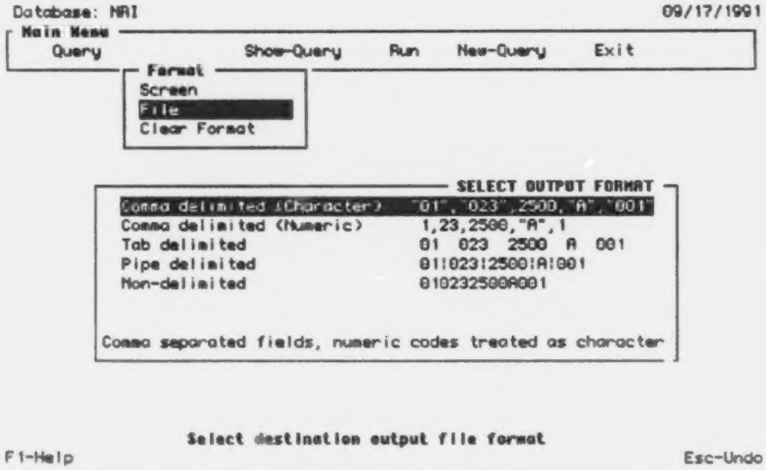
SCREEN The screen display is designed to offer a snapshot of the data for the user to decide if they are retrieving what they intended. INFeRS was designed to extract subsets from a large dataset. If a system was designed to retrieve mostly summary numbers, a full screen reporting module would need to be designed.



```
 1: 48 039       15: 2        30: 000          45: 99        60: 1
 2: 0250974      16: 2        31: 01           46: 99        61: 02
 3: 1            17: 2        32: 32           47: 99        62: 1
 4: 7035         18: 1        33: 352          48: 00        63: 0
 5: 150A         19: 1        34: 000300       49: 0         64: 0
 6: 12040205     20: A        35: 100          50: 0         65: 0
 7: 000026       21: A        36: 0125         51: 0         66: 0
 8: 1            22: 006      37: 0002         52: 0         67: 0
9a: 2            23: 11       38: 0            53: 0         68: 0
9b: W            24: 006      39: 000003       54: 0000      69:
10: 5            25: 222      40: 000000000    55: 9999      70: 000
11: 1            26: 222      41: 000000       56: 0310      71: 0
12: 1            27: 2        42: 000000000    57: 0400      72: 000
13: 2            28: 449      43: 99           58: 1200      73: 000
14: 7            29: 000      44: 99           59: 0650      74: 0
                                                             75: 0
Soils-5: TX0758   C   0001A010101211111111111111111111111111

Record #: 1


Press RETURN to continue ... <a> to abort ...
```

figure 26 Screen Display, one record

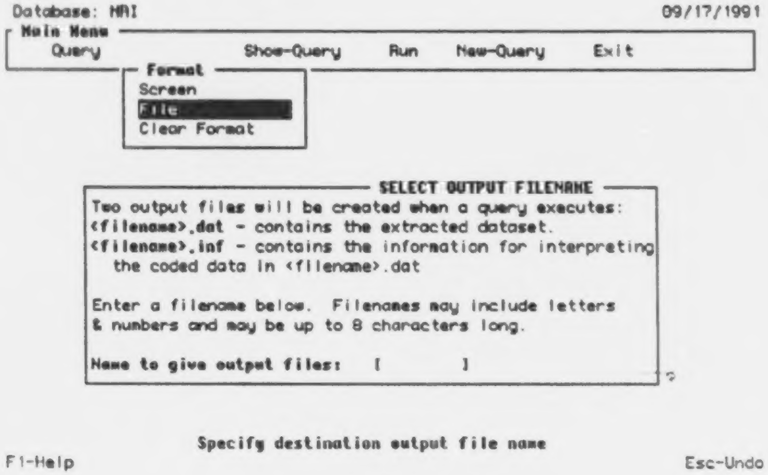
FILE The file format options allows users to specify one of five formats for the output file.



```
Database: MAI                                              09/17/1991
┌ Main Menu ──────────────────────────────────────────────────────────┐
    Query                  Show-Query    Run    New-Query    Exit
└──────────────── Format ─────────────────────────────────────────────┘
                 Screen
                 File
                 Clear Format

                                          SELECT OUTPUT FORMAT
   Comma delimited (Character)     "01","023",2500,"A","001"
   Comma delimited (Numeric)       1,23,2500,"A",1
   Tab delimited                   01  023  2500  A  001
   Pipe delimited                  01|023|2500|A|001
   Non-delimited                   010232500A001

  Comma separated fields, numeric codes treated as character


              Select destination output file format
F1-Help                                                    Esc-Undo
```

figure 27 File Format Window

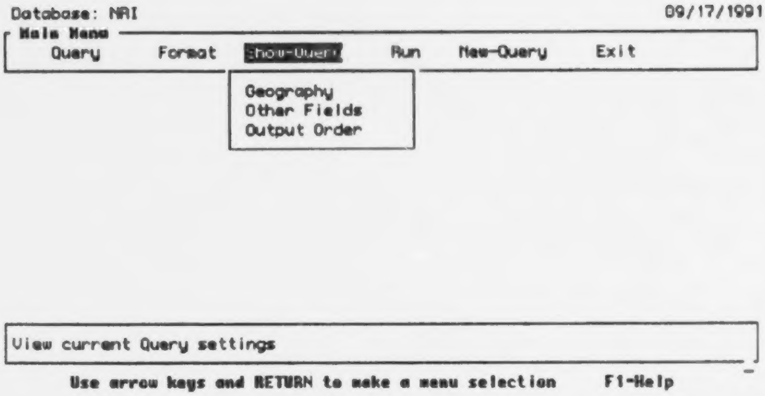
Once they have highlighted and selected a format they are asked to type in a file name prefix. INFeRS automatically assigns the suffixes .inf and .dat to the information/documentation and data files created.



```
Database: MAI                                              09/17/1991
┌ Main Menu ──────────────────────────────────────────────────────────┐
    Query                  Show-Query    Run    New-Query    Exit
└──────────────── Format ─────────────────────────────────────────────┘
                 Screen
                 File
                 Clear Format

                                   SELECT OUTPUT FILENAME
  Two output files will be created when a query executes:
  <filename>.dat - contains the extracted dataset.
  <filename>.inf - contains the information for interpreting
    the coded data in <filename>.dat

  Enter a filename below.  Filenames may include letters
  & numbers and may be up to 8 characters long.

  Name to give output files:   [        ]


              Specify destination output file name
F1-Help                                                    Esc-Undo
```

figure 28 Filename Window

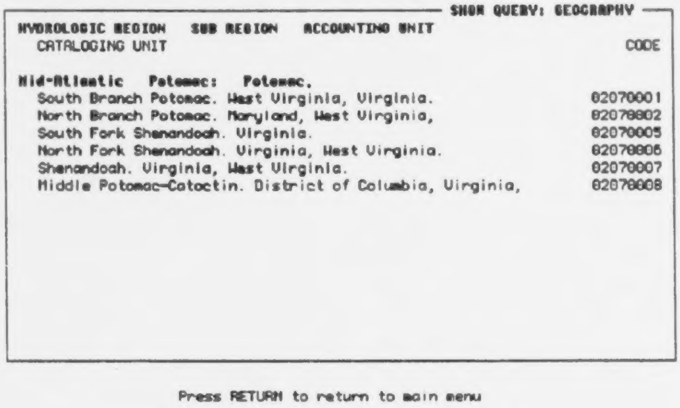
Show Query



```
Database: NRI                                                   09/17/1991
┌ Main Menu ──────────────────────────────────────────────────────────────┐
│   Query       Format     Show-Query     Run     New-Query     Exit      │
└─────────────────────────────────────────────────────────────────────────┘
                        Geography
                        Other Fields
                        Output Order

View current Query settings
        Use arrow keys and RETURN to make a menu selection      F1-Help
```

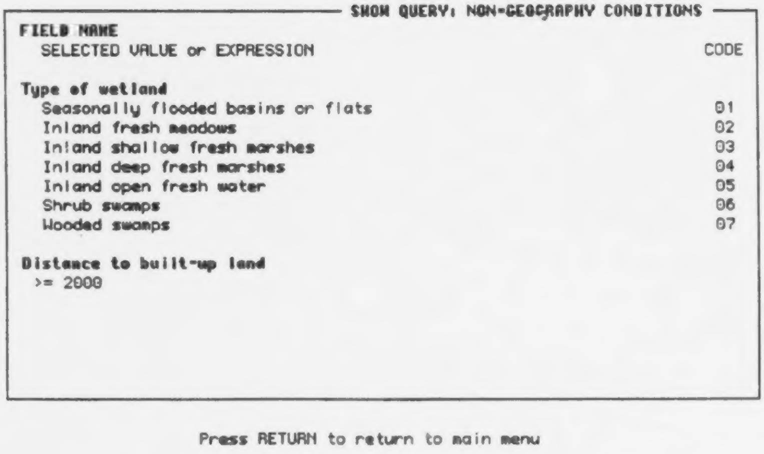
figure 29 Show Query Options

The Show Query function is optional. It allows a user to review the search that has been constructed, or as it is being constructed. They view their geographic conditions and/or other field conditions.



```
                                                  SHOW QUERY: GEOGRAPHY
HYDROLOGIC REGION   SUB REGION   ACCOUNTING UNIT
  CATALOGING UNIT                                                   CODE

Mid-Atlantic   Potomac:   Potomac.
  South Branch Potomac. West Virginia, Virginia.                02070001
  North Branch Potomac. Maryland, West Virginia,                02070002
  South Fork Shenandoah. Virginia.                              02070005
  North Fork Shenandoah. Virginia, West Virginia.               02070006
  Shenandoah. Virginia, West Virginia.                          02070007
  Middle Potomac-Catoctin. District of Columbia, Virginia,      02070008

                   Press RETURN to return to main menu
```

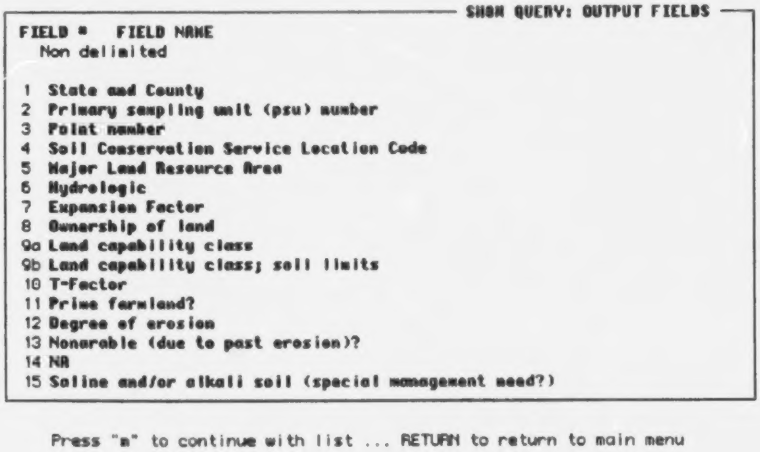
figure 30 Geography Show Query Screen



```
                              SHOW QUERY: NON-GEOGRAPHY CONDITIONS
FIELD NAME
  SELECTED VALUE or EXPRESSION                                      CODE

Type of wetland
  Seasonally flooded basins or flats                                  01
  Inland fresh meadows                                                02
  Inland shallow fresh marshes                                        03
  Inland deep fresh marshes                                           04
  Inland open fresh water                                             05
  Shrub swamps                                                        06
  Wooded swamps                                                       07

Distance to built-up land
 >= 2000

                   Press RETURN to return to main menu
```

figure 31 Other Fields Show Query Screen

Show Query also displays the fields selected for output, in the order they will be output.

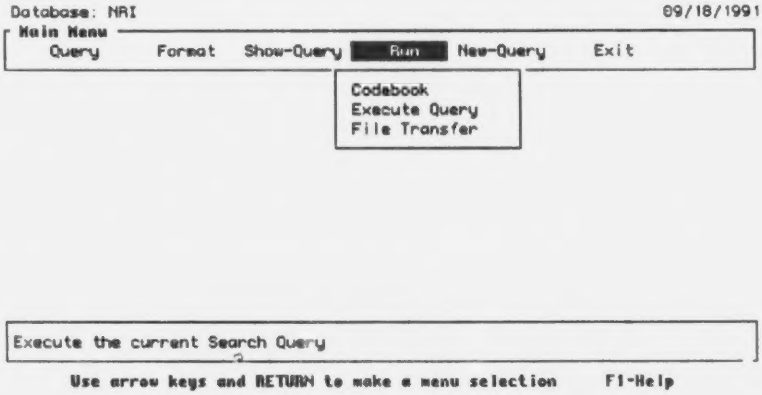


```
FIELD #   FIELD NAME                              SHOW QUERY: OUTPUT FIELDS
  Non delimited

 1  State and County
 2  Primary sampling unit (psu) number
 3  Point number
 4  Soil Conservation Service Location Code
 5  Major Land Resource Area
 6  Hydrologic
 7  Expansion Factor
 8  Ownership of land
 9a Land capability class
 9b Land capability class; soil limits
10  T-Factor
11  Prime farmland?
12  Degree of erosion
13  Nonarable (due to past erosion)?
14  NA
15  Saline and/or alkali soil (special management need?)

      Press "n" to continue with list ... RETURN to return to main menu
```

figure 32 Output Order Show Query Screen

Run

Once all the definitions and selections have been made the search can be run.



Execute the current Search Query

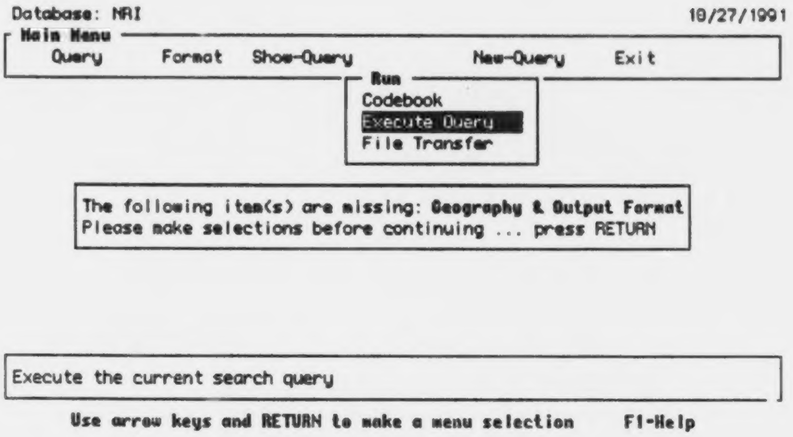
Use arrow keys and RETURN to make a menu selection

figure 33 Run Search Menu Options

CODEBOOK The user can opt to receive the full documentation (even information on fields that are not included in the output) by selecting the codebook option from the menu. If they do the selection is flagged with a >.

EXECUTE QUERY Once the Execute Query selection is made the system takes over, assembles the SQL statement, sends it to the database, and formats the resulting records and associated documentation. Details of this procedure are given in the System Section.

If part of the search definition is missing, INFeRS alerts the user.



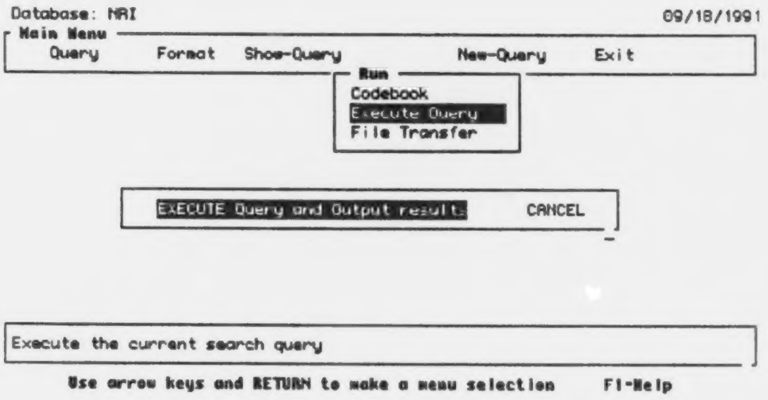
Database: NRI
10/27/1991

The following item(s) are missing: **Geography & Output Format**
Please make selections before continuing ... press RETURN

Execute the current search query

figure 34 Search Incomplete Message

Then the user is given an opportunity to change their mind.

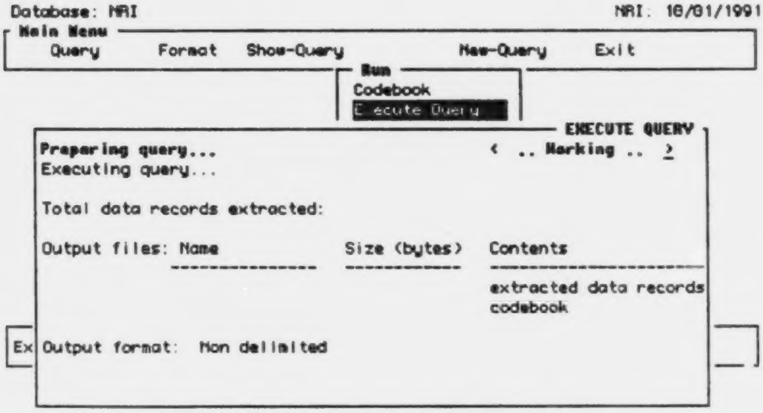


Database: NRI
09/10/1991

Execute the current search query

figure 35 Execute Query Confirmation Window

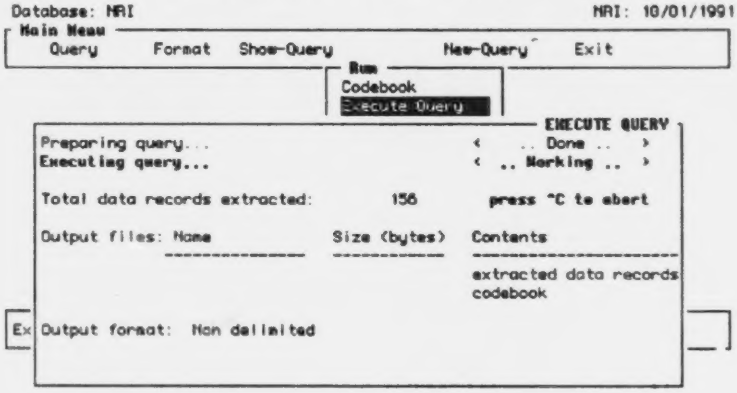
INFeRS displays the search in progress through a search tracking window that indicates when the query (SQL statement) has been prepared, and when the searching is in progress.



```
Database: NRI                                          NRI: 10/01/1991
Main Menu
  Query     Format   Show-Query            New-Query     Exit
                                Run
                                Codebook
                                Execute Query
                                                        EXECUTE QUERY
Preparing query...                                  < .. Working .. >
Executing query...

Total data records extracted:

Output files: Name              Size (bytes)     Contents
              ---------------   ------------     ----------------------
                                                  extracted data records
                                                  codebook

Output format:   Non delimited
```

figure 36 Search Window, 'preparing query'

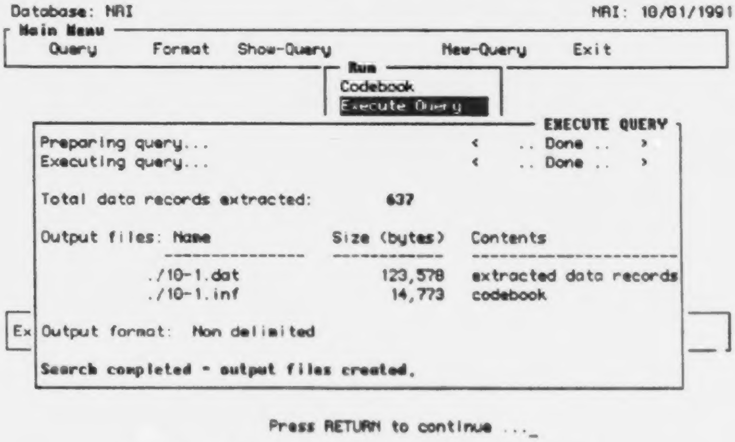
The number of records extracted is displayed as the system searches. If the set becomes to large for the user to manage they can break the search anytime in this process by pressing 'control' C.



```
Database: NRI                                          NRI: 10/01/1991
Main Menu
  Query     Format   Show-Query            New-Query     Exit
                                Run
                                Codebook
                                Execute Query
                                                        EXECUTE QUERY
Preparing query...                                  <   .. Done ..   >
Executing query...                                  < .. Working ..  >

Total data records extracted:          156          press ^C to abort

Output files: Name              Size (bytes)     Contents
              ---------------   ------------     ----------------------
                                                  extracted data records
                                                  codebook

Output format:   Non delimited
```

figure 37 Search Window, record number display

When the search is complete the window shows the sizes and names of the files created.

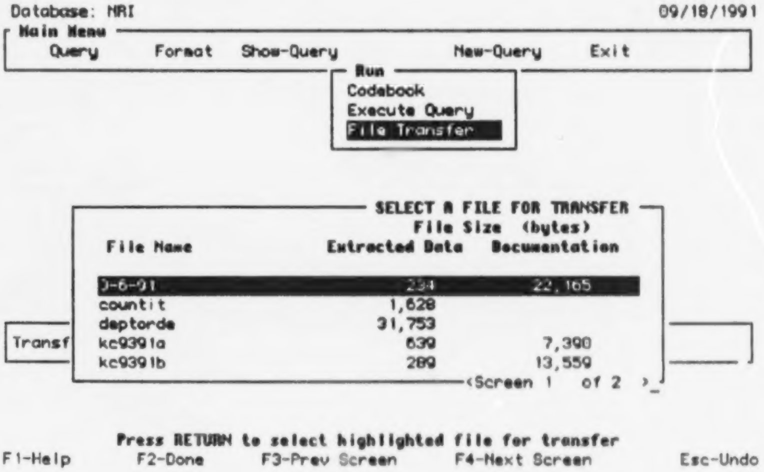


```
Database: NRI                                          NRI: 10/01/1991
Main Menu
  Query     Format   Show-Query            New-Query     Exit
                                Run
                                Codebook
                                Execute Query
                                                        EXECUTE QUERY
Preparing query...                                  <   .. Done ..   >
Executing query...                                  <   .. Done ..   >

Total data records extracted:          637

Output files: Name              Size (bytes)     Contents
              ---------------   ------------     ----------------------
                ./10-1.dat          123,578      extracted data records
                ./10-1.inf           14,773      codebook

Output format:   Non delimited

Search completed - output files created.

                    Press RETURN to continue ..._
```

figure 38 Search Window, search completed

FILE TRANSFER Once the search is complete the user is free to revise the search, start a new search, or transfer the output files to his or her local system. All files generated in a session (until logoff the system) are stored for downloading.

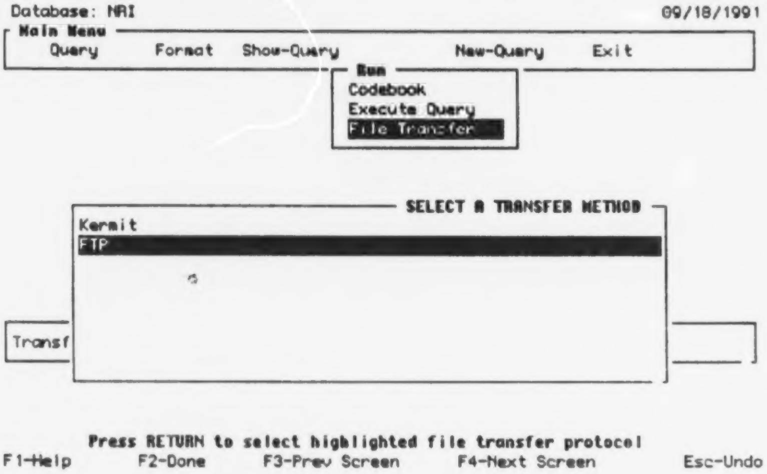
When File Transfer is selected from the Run Menu the user sees a list of the files INFeRS has generated for output.



```
                              SELECT A FILE FOR TRANSFER
                                     File Size  (bytes)
 File Name                  Extracted Data   Documentation

 0-6-91                           234           22,165
 countit                        1,628
 deptorde                      31,753
 kc9391a                          639            7,390
 kc9391b                          289           13,559
                                      <Screen 1   of 2  >

        Press RETURN to select highlighted file for transfer
```

figure 39 Select a File for Transfer Window

They select the file (both the .inf and .dat files) for transfer and press Enter. They are asked to select a transfer method. At present INFeRS supports Kermit (modem type) and ftp (high speed) modes.

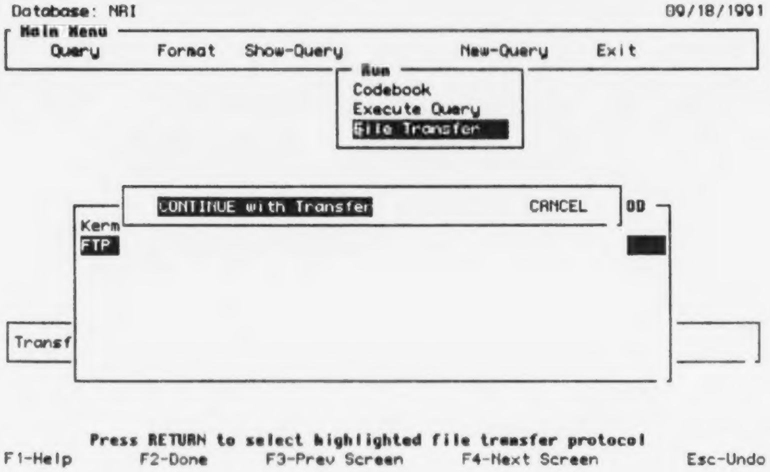


```
                              SELECT A TRANSFER METHOD
 Kermit
 FTP

        Press RETURN to select highlighted file transfer protocol
```

figure 40 Select a Transfer Method Window

At this point 'control' of the process is moved outside of Informix, so users are given the first of two opportunities to cancel the process.

BEST COPY AVAILABLE

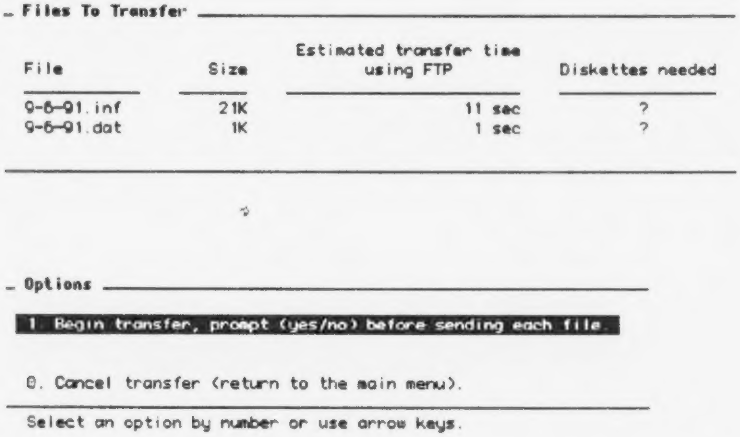


```
Database: NRI                                                    09/18/1991
Main Menu
   Query      Format    Show-Query              New-Query      Exit
                                      Run
                                      Codebook
                                      Execute Query
                                      File Transfer

                CONTINUE with Transfer                  CANCEL    OD
          Kerm
          FTP

Transf

              Press RETURN to select highlighted file transfer protocol
F1-Help      F2-Done      F3-Prev Screen      F4-Next Screen            Esc-Undo
```

figure 41 Continue with Transfer Window

A generic file transfer module is then invoked. The first screen of that module gives the patron information on the size of the files selected for transfer and the estimated transfer file. Large files could take from several minutes to hours to transfer. In a future version of this module the user will be asked what sort of disk they are using to store the file before this screen is displayed. Then the fourth column in the table will display number of disks required to hold the file.

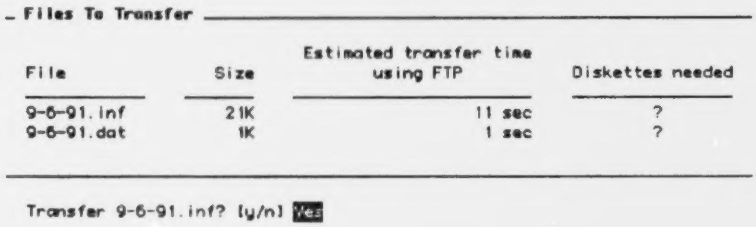
If the estimated transfer time is unacceptable to the user they can cancel the process. Otherwise they select the default option: continue.



```
_ Files To Transfer ____________________________________________________

                            Estimated transfer time
File              Size            using FTP              Diskettes needed
______________    ____    _______________________      ________________
9-6-91.inf         21K                     11 sec                ?
9-6-91.dat          1K                      1 sec                ?
________________________________________________________________________



_ Options _____________________________________________________

 1. Begin transfer, prompt (yes/no) before sending each file.

 0. Cancel transfer (return to the main menu).
_______________________________________________________________
 Select an option by number or use arrow keys.
```

figure 42 File Transfer, continue option

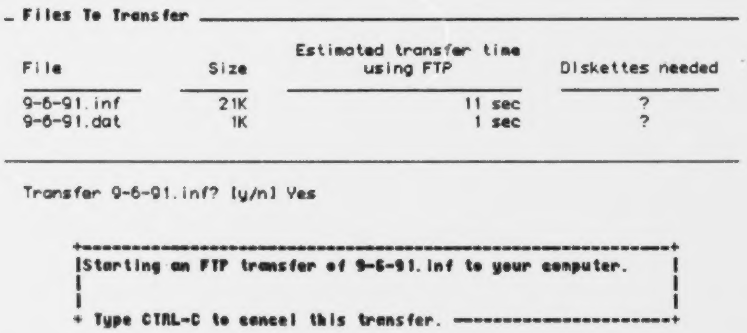
The window then changes to display the file names selected for transfer, with a yes/no option for each file.



```
_ Files To Transfer ______________________________________________

                                 Estimated transfer time
File            Size                    using FTP             Diskettes needed
_____________   ________   ______________________________   __________________
9-6-91.inf       21K                              11 sec             ?
9-6-91.dat        1K                               1 sec             ?
_______________________________________________________________________________

Transfer 9-6-91.inf? [y/n] Yes
```

figure 43 File Transfer, individual file selection

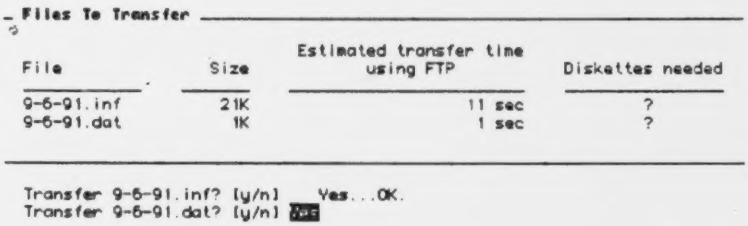
The actual transfer is then initiated, the user can interrupt the process by pressing 'control' C.



```
_ Files To Transfer ______________________________________________

                                 Estimated transfer time
File            Size                    using FTP             Diskettes needed
_____________   ________   ______________________________   __________________
9-6-91.inf       21K                              11 sec             ?
9-6-91.dat        1K                               1 sec             ?
_______________________________________________________________________________

Transfer 9-6-91.inf? [y/n] Yes

      +---------------------------------------------------------------+
      |Starting an FTP transfer of 9-6-91.inf to your computer.       |
      |                                                               |
      |                                                               |
      + Type CTRL-C to cancel this transfer. -------------------------+
```

figure 44 File Transfer, transfer process

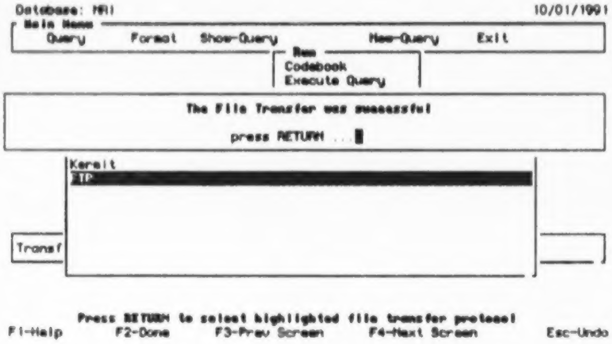
If the transfer is successful an "OK" message appears next to the file name and the user accepts the next file to be transferred



```
_ Files To Transfer ______________________________________________

                                 Estimated transfer time
File            Size                    using FTP             Diskettes needed
_____________   ________   ______________________________   __________________
9-6-91.inf       21K                              11 sec             ?
9-6-91.dat        1K                               1 sec             ?
_______________________________________________________________________________

Transfer 9-6-91.inf? [y/n]    Yes...OK.
Transfer 9-6-91.dat? [y/n] Yes
```

figure 45 File Transfer, transfer process success

When all the selected files have been transferred control is returned to INFeRS and a success message is displayed.



The File Transfer was successful

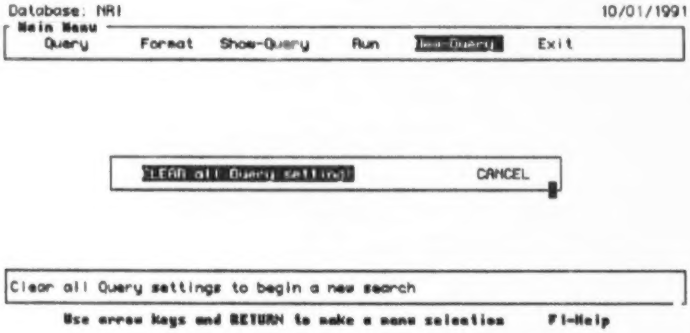
press RETURN ...

Press RETURN to select highlighted file transfer protocol

figure 46 Successful Transfer Window

New Query

The definitions set in the geography and other fields indexes and file formats and names are retained until they are individually 'erased' by the user, or the user logs off the system, or they use the New Query function to erase all the definitions.



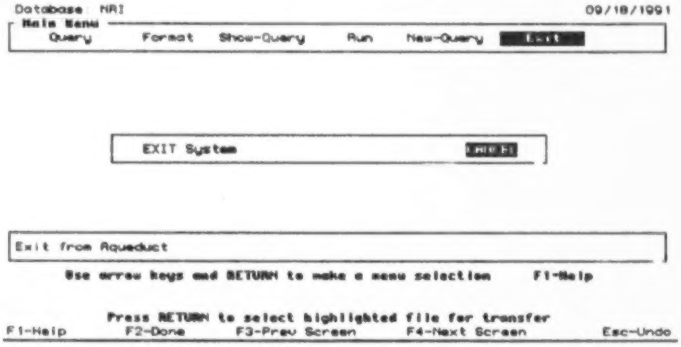
Clear all Query settings to begin a new search

Use arrow keys and RETURN to make a menu selection

figure 47 Clear all Query Settings Window

Exit

When the user has completed a session they log off INFeRS. Once again they are given the opportunity to rescind their order, and since the default option is Cancel the exit, they must actively select the Exit option.



Exit from Aqueduct

Use arrow keys and RETURN to make a menu selection

Press RETURN to select highlighted file for transfer

figure 48 Exit Window

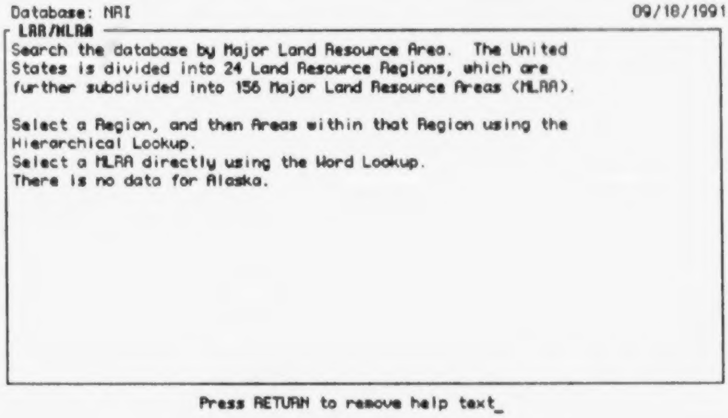
Helps

Help on the system is provided at two points: online, and in the information file part of the output.

Online help

All online help is context specific and retrieved when F1 is pressed. We divided the online help into different types: A description of the field, or option highlighted at the time Help was requested; a description of the screen open at the time; and the general navigation and actions possible from that screen. Users are given a menu of choices when they press F1.

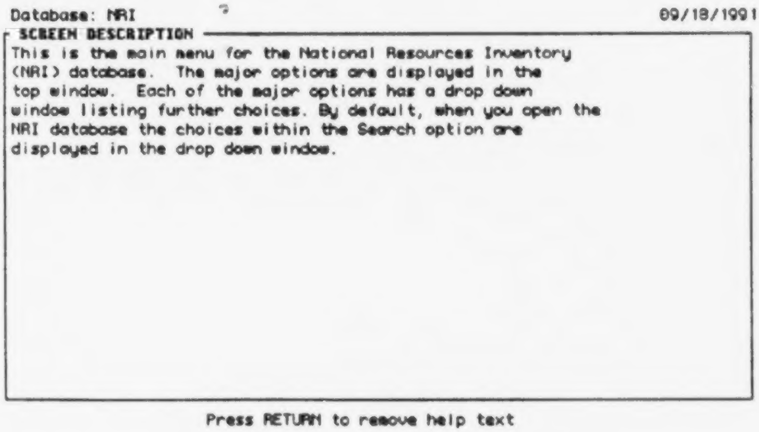
FIELD DESCRIPTION If they select 'description of the field or option line' they are given further definitions or descriptions of the field, and a list of the coded values and their names (or a sample, if the list is over a screen.)



```
Database: NRI                                              09/18/1991
 LAR/MLRA
Search the database by Major Land Resource Area.  The United
States is divided into 24 Land Resource Regions, which are
further subdivided into 156 Major Land Resource Areas (MLRA).

Select a Region, and then Areas within that Region using the
Hierarchical Lookup.
Select a MLRA directly using the Word Lookup.
There is no data for Alaska.

                    Press RETURN to remove help text_
```

figure 49 Land Resource/MLRA description

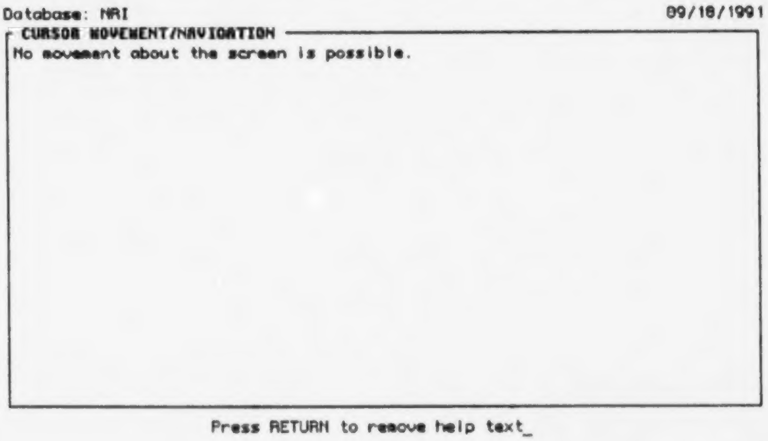
SCREEN DESCRIPTION The screen description is basically a navigation aid, reminding the user of where they are in the search sequence.



```
Database: NRI                                              09/18/1991
 SCREEN DESCRIPTION
This is the main menu for the National Resources Inventory
(NRI) database.  The major options are displayed in the
top window.  Each of the major options has a drop down
window listing further choices. By default, when you open the
NRI database the choices within the Search option are
displayed in the drop down window.

                    Press RETURN to remove help text
```

figure 50 Screen Description, Main Menu

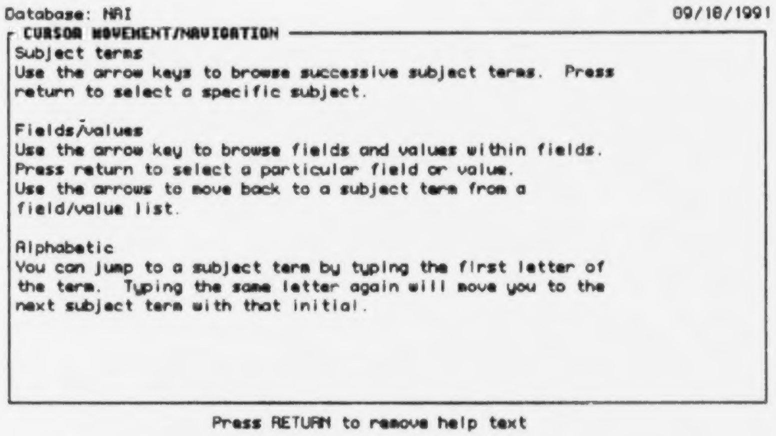
NAVIGATION/CURSOR MOVEMENT Cursor movements range from quite simple:



```
Database: NRI                                              09/18/1991
┌ CURSOR MOVEMENT/NAVIGATION ──────────────────────────────────────┐
No movement about the screen is possible.

                    Press RETURN to remove help text_
```

figure 51 Cursor Movement, in numeric condition fill-in window

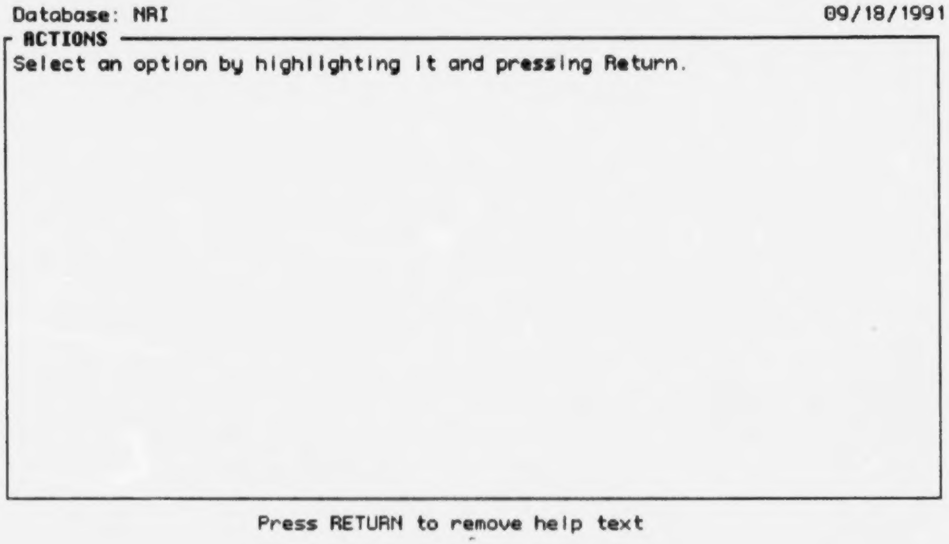
to quite complex:



```
Database: NRI                                              09/18/1991
┌ CURSOR MOVEMENT/NAVIGATION ──────────────────────────────────────┐
Subject terms
Use the arrow keys to browse successive subject terms.  Press
return to select a specific subject.

Fields/values
Use the arrow key to browse fields and values within fields.
Press return to select a particular field or value.
Use the arrows to move back to a subject term from a
field/value list.

Alphabetic
You can jump to a subject term by typing the first letter of
the term.  Typing the same letter again will move you to the
next subject term with that initial.

                    Press RETURN to remove help text
```

figure 52 Cursor Movement, Subject Index

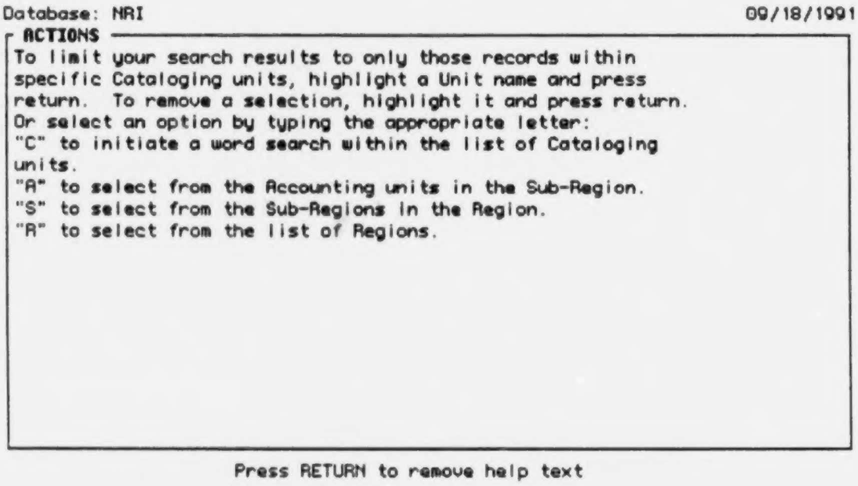
ACTION The Action screen gives users function/sequence assistance, alerting them to the alternatives and the outcomes of their selections. They can also be quite simple:



```
Database: NRI                                               09/18/1991
ACTIONS
Select an option by highlighting it and pressing Return.

Press RETURN to remove help text
```

figure 53 Actions, Main Menu

or quite complex:



```
Database: NRI                                         09/18/1991
ACTIONS
To limit your search results to only those records within
specific Cataloging units, highlight a Unit name and press
return.  To remove a selection, highlight it and press return.
Or select an option by typing the appropriate letter:
"C" to initiate a word search within the list of Cataloging
units.
"A" to select from the Accounting units in the Sub-Region.
"S" to select from the Sub-Regions in the Region.
"R" to select from the list of Regions.

Press RETURN to remove help text
```

figure 54 Actions, Hydrologic Units Window

All the Helps are coded into the text lookup tables, so their content can be altered without altering program code.

Output helps

Unless the user selects 'Codebook' from the Run Menu they are given only the information they need to decipher their subset. If they select Codebook they are given the complete codebook for the file. We gave users the option for an abbreviated version since they might be creating many subsets and would not need a codebook for each one. A sample information file is are included as Appendix D.

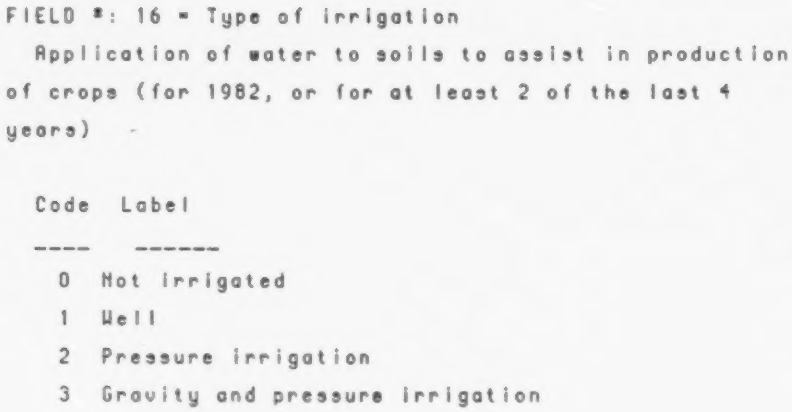
In the information file users are first warned about the most likely misuse of the file: the reliability of analysing the data at different geographic unit-levels. We then give them citations for further information about the file, and several of the more complex coded fields, such as the Hydrologic Units and the Major Land Resource Areas. Then we list the source for further information about the data, and the use of INFeRS.

The information specific to the output follows:

—the fields output, in order of output

—the search conditions set, starting with geography, and then other conditions (if they were set).

Finally the codes for all the output fields are listed. Each field is listed by number (corresponding to the initial list of fields to be output), and by name of field with the code for the value and the description of the code.



```
FIELD #: 16 = Type of irrigation
  Application of water to soils to assist in production
of crops (for 1982, or for at least 2 of the last 4
years)

  Code  Label
  ----   ------
    0  Not irrigated
    1  Well
    2  Pressure irrigation
    3  Gravity and pressure irrigation
```

figure 55 Sample Field/codebook entry

Conclusion

Still To Do

Testing

We have tested INFeRS to see if the interface is intelligible to the user. We now want to test the whole system to determine who of our users will need the functions provided by INFeRS. Our audience of users range from data experts to data novices, and they may be computing experts or computing novices. We want to answer two questions for the range of potential users of the system.

—How much of the system *can* these different users use? Can they take advantage of all of the functions of the system?

—How much of the system *will* these different users use? How much do they actually need for their work?

In order to answer these questions, test subjects will do a sequence of five searches designed to use successively sophisticated functions of INFeRS. We will measure how much of the system each 'type' of user can use. Then we will ask them which of the search scenarios most closely resembles what they would do for their research or coursework. We will consider INFeRS 'successful' if a user can accomplish the level of tasks they would need. For example, if a power-user is able to accomplish all five of our scenarios, but would usually use up to the third level, INFeRS would be useful. Or, if a computer and data novice is only able to complete the first-level search, but would only ever need such a search, INFeRS serves its purpose.

Enhancements

We have a long list of small fixes and enhancements for the system. That list comes from our testing and evaluation of the system to date. We also have a smaller list of more involved enhancements that was compiled during the design and creation of INFeRS. In no particular order they are discussed below.

SAS/SPSS OUTPUT. It would be useful to include this as an additional option in the file format level. A user who intends to load the extracted data into SAS will have to generate a small SAS job with the locations, and labels, of the data. It would be fairly straighforward to create that job in INFeRS and output it as an additional file for the user.

DOCUMENTATION LIBRARY. Instead of telling the user that additional information about the file is available in the various publications listed in the information file it would be nice to allow the user to transfer that information from INFeRS to their computer. This option could be included in the RUN menu where we currently give a user the option to output the full codebook. If the option read DOCUMENTATION instead of CODEBOOK, we could open a window with a list of documentation available for downloading.

SUMMARY STATISTICS FILE. All the files we have planned to load are fairly large files where an automated extraction system is a definite advantage. It would be interesting to load a large number of small files of summary statistics, such as tables from the *Statistical Abstract*. These files would appeal to people who just need a few summary numbers for a report or calculation.

CROSS FILE SEARCHING. We made a deliberate decision not to merge files. (See next section: Summary of Major Issues for details.) But, if a user were searching for information about a particular county in several files, they would have to select that county in each file. We want to create a function that will allow users to 'carry-over' certain types of selections. In our discussion we decided that they could carry-over definitions based on standards, such as the Federal Information Processing Standards (FIPS). For databases using FIPS codes for states and counties, users would be able to save those selections for use on another dataset that also adhered to FIPS state/county codes. The U.S. Geological Survey Hydrologic Codes are also a FIP Standard. Even using FIPS codes could create problems because the definitions do change, some states have added, merged, and deleted county codes over the years.

To allow any other carry-over variables would be quite dangerous. Even such seemingly standard values such as year can vary from database to database. The Crop Reporting Board years are harvest years, not calendar years. Other databases might use budget years. A 'year' saved from one search and run against another file could retrieve erroneous data.

DIRECT SQL SEARCHING. In some cases INFeRS would not be the quickest way to search the database. In some circumstances a patron might know how to formulate a SQL statement and would find it easier to run SQL searches again the database directly. We want to create a function to allow those sorts of direct searches.

GRAPHIC INTERFACE. There are microcomputer applications that will run against mainframe databases. We would like to create a distributed interface to run on the user's local system. That interface could take advantage of the high-end capabilities of each system, instead of both machines dropping to the VT100 terminal interface.

CROSS-FILE INDEX. Currently a user has to open each file to see the values/variables available in the datafile. At a minimum they should be able to search a merged index of all the files to determine which files have information of interest. This meta-catalog could be quite complex, allowing users to tag several files to be searched, automatically saving the allowable cross-file variables the user defines, and moving on to each file in turn.

CALCULATION MODULE. In some of the files certain calculations of the data would be needed regularly. For example multiplying the expansion factor for each sample point record of the National Resources Inventory for data collected at the Major Land Resource Unit Level. In the Health and Nutrition Examination Survey data the weighting for each record is crucial. Each individual record represents a certain number of people in the United States. In order to draw conclusions on percentages in any particular category the user would have to weight each record in the data subset.

The system would be more useful to the 'power-users' if modules were added to perform those semi-standard data calculations. The inherent danger would be if those modules were used by naive users who did not fully understand what was going on, they could create nonsense numbers.

Summary of Major Issues

Several issues came to our attention repeatedly as we researched, designed, and developed INFeRS. We read about them, heard about them from users, and encountered them creating the system.

Merging data

One of our original thoughts was to provide users with seamless access to data from different sources. We thought it would be marvellous to create a system where a user could retrieve all the data on 'corn' in the system. In practicality we discovered most of our users want to know where their data are coming from, and want them kept separate. To quote one soil scientist "I'm only going into court with data from the U.S.G.S. and the SCS." Time after time our interviewees emphasized control over data source. Controlling the data extracted is one of the ways researchers can ensure the quality of their analysis. Our compromise design would be to allow cross file searching for standard elements. Another possibility would be to allow users to retrieve all the subsets with 'corn' in the values or variables, but also to provide definitions for how corn was defined in each case (white, yellow, sweet, etc.)

Heavy dataset users vs. tangential data users

One of the problems in designing INFeRS was meeting our goal of providing access to the file for both the data expert and the patron who just needed a number or two for a paper. The needs of these users are quite different, essentially they are at two ends of a continuum. We concentrated our design on protecting the tangential user as much as possible without compromising the ability of the power-user to retrieve data. For example, in the Crop Estimates file INFeRS allows a user to select several commodities, and specify numeric conditions for other fields. A user could select apples and wheat, and specify a production level. But that number would be run against apple production in pounds and wheat production in bushels, in other words a nonsense search. INFeRS warns the user about this danger in a series of information screens and allows the user to restructure the search. An experienced user would not make that kind of mistake. But we could not design the system assuming experienced users.

One Interface

We designed an interface that could be used for a wide range of files with wildly different internal structures. The interface is process-defined, rather than data-structure-defined. All of the idiosyncrasies of the individual dataset are accommodated in the first, Search, window. Although both datasets loaded to date contain geographic elements, the system can handle other types of files. We could load a genetic sequence database into the format created, or a database of physical constants. The important thing is the consistency the user will have in searching any of the files we might load. INFeRS can compensate for most data vagaries in the individual files, the user does not have to discover them.

Standards

INFeRS was created to deal with multiple datasets encoded and documented with no internal standards, or differing standards. We encountered problems with standards in other aspects of the project: system access and file transfer.

VT100. We discovered that VT100 is not a standard. It defines most keys, but others (such as arrow keys) are not included in that standard definition. A system that supports VT100 may not support arrow keys. We tested our system against most of the software packages used by our patrons and discovered that several of them would not allow the searcher to use the arrow, a major liability for INFeRS searching. Our temporary solution is to provide users with public-access programs that do support the full range of key strokes needed to search INFeRS. In our sign-on literature we describe the settings users will need to have. But access should be as simple as telling a user to use VT100 terminal emulation.

FILE TRANSFER. Computer to computer communication is still in a state of anarchy. Standards and common protocols are becoming more common, but we still found a wide range of communications patterns amongst our users. Some have no access to other computers, some users have modems, others have slow or fast direct connections to the university network. We had to design a file transfer routine that accounted for all the common communication profiles of our users. When transfer standards become commonplace users will have an easier time moving data from systems such as INFeRS to their local computer.

Resources

Interface/database creation is an extremely time consuming affair. This is not news. In hindsight, we created our system with too few people. We could have created several of the enhancements listed above if we had the 'ideal' project mix of skills and manpower hours. Since we were trying to demonstrate the feasibility of a system we concentrated on system design. When we realized the complexity of the National Resources Inventory file, we made a decision to concentrate on what we could learn from loading that file, rather designing a lesser system in order to spend our time loading additional files.

We have learned much about how a library can create interfaces to databases. Libraries bring a philosophy, or mindset, to the creation of databases that traditional creators of such databases lack. We are uniquely aware of the great variety of uses of data. The drawback to this attitude is a library tends to try and create a system that is all things to all users, which is an impossible task.

If we were to repeat our project we would want three sets of staff: one for system creation, one for adding new files to an existing system, and one for maintenance of a system.

SYSTEM CREATION — nine months

Administrator, user advocate—This person would determine system specifications and oversee the progress of the project. A 50% assignment to the project would suffice.

Database designer/programmer—see next position

Interface designer/programmer—To create a system at least two people with programming skills and either database design or interface skills are needed. Both of those positions should be full time on the project.

Datafile expert, statistician—This person would run the statistical analyses on the file, research how the data were collected, and retrieve all the relevant documentation for the file. They would create any indexes for the file. This person would need to work full-time for 4-8 weeks on each file in order to 'clean it up' to the level we suggest in the Data Section.

ADDING A NEW FILE — one to two months

Administrator—decide on the file, determine any additional system specifications over the original system, and oversee the project.

Datafile expert, statistician—determine if the data structure dictates any changes in the database or interface, decide how it should be merged into the system, and create any indexes.

Database designer/programmer—modify the system to incorporate the new file.

MAINTENANCE — ongoing

Programmer—to maintain system, debug program, and update any ongoing datasets.

From Research and Demonstration to Reality

Would we do it again? We do not need to, we have a working system. The question becomes: Will we maintain our system? Yes, because we can and we should. INFeRS is simply an expansion of the current activities of our technical and public services. Just as libraries are creating online table-of-contents services for their collections (e.g., CARL Uncover) we are creating an online table-of-contents to our numeric computer files. Libraries have created online catalogs with standard levels of bibliographic description to allow users to determine whether or not they want to look at a particular volume. INFeRS is a system that allows users to determine whether or not they want a particular piece of a datafile. Circulation departments exist to assist patrons in using the collection. The extraction and File Transfer modules in INFeRS was created to serve the same purpose.

Of course, given the labor intensive nature of this system, we can only provide this level of service for some of our data. But as we argued in our original proposal in 1988, just as libraries have shared cataloging efforts and print resources, we should share database development and computerized numeric data. INFeRS is available to anyone with access to the Internet. No other library should have to create a system to access the data we have on our computers. In the future it would be nice to have minimum standards for what a numeric interface should do, such as: extract to the variable level; allow numeric or equivalent search limiting; supply adequate codebook information; and transfer the data in some standard formats.

Data (e.g., information) is inherently confusing, analysis (e.g., scholarship) is our only hope of making it sensible. Numeric interfaces are an attempt to simplify the extraction step so the scholar can concentrate on the real job of analysis. So, in building INFeRS we have only done what libraries have always done.

Appendices

Appendix A. Request for Proposal: Minicomputer

Appendix B. Request for Proposal: Database Management Software

Appendix C. Sample Data Output, comma delimited

Appendix D. Full Codebook

Appendix E. National Level Presentations of INFeRS

Appendix F: Handouts from presentations of INFeRS

Appendix A: Request for Proposal — Hardware

Mann Library Information Server Procurement

Request for Proposal

1. Introduction: Mann Library, the library to the College of Agriculture and Life Sciences and the College of Human Ecology at Cornell University, seeks bids on a computer and associated hardware and software systems. Initially, the library will use this computer to develop an interactive system that permits researchers to identify and extract subsets from heavily used data files. The system, funded by a grant from the U.S. Department of Education, involves mounting large files of statistical and census data on online storage devices under the control of database management software. Library information specialists and technical staff will then develop a search-and-retrieval interface to this data resource which will assist researchers in extracting data sets from the central files. (A more extensive description of this project follows this document as Attachment A).

In the future, the library will also move its "scholarly information system"--currently running under VMS on a MicroVax computer--to the new computer. The development of a scholarly information system at Cornell is a joint project of Mann Library and Cornell Information Technologies. The system now provides access to two major bibliographic files--AGRICOLA and subsets of the BIOSIS Previews database--through BRS/Search, a textual database management package licensed from BRS Information Technologies of Latham, New York. The system will be used on an experimental basis during spring semester 1989 by Cornell researchers in the departments of Entomology, Genetics, and Nutritional Science. Mann Library and CIT will later provide this information service to a broader clientele.

A second track of the scholarly information system project involves mounting files of digitized full-text information on optical disk for distribution to workstations in scholar's offices through the Cornell campus network. (A more extensive description of this project follows this document as Attachment B).

2. Guidelines for proposal preparation:

2.1 This request for proposal divides the system description into a number of discrete areas. Interested vendors should prepare their bid according to this breakdown. Please specify the hardware and software items that you are proposing to supply in each area, the list price of each item, and the price at which you are bidding the item.

If the system you wish to propose cannot be divided into the specified areas, please explain.

2.2 In those areas where the vendor's personnel or agents will need to carry out installation or configuration activities, please specify the nature and scope of these tasks. Also specify any charges that will result for these services, presenting: job title of individual responsible for work, estimated hours of work, and charge for completion of activity. Please specify any areas in which charges for installation services are not fixed in advance.

2.3 In those areas where Mann Library or Cornell Information Technologies will have to provide significant installation services to make a component of the system operational, please describe your understanding of the nature of the work that Cornell must undertake. Provide an estimate of the number of hours of labor, broken down by function (i.e., "Systems engineer--2 hours; network technician--3 hours"), that you believe this work will require. Note: These estimates are for the information of Mann Library and are not binding on the vendor.

2.4 Note that Mann Library is a unit of the statutory colleges at Cornell University and is thus eligible for all pertinent New York State discounts and advantages.

2.5 Note that the nature of the Department of Education grant should qualify this system for consideration as a research application and qualify Mann Library for any pertinent vendor discounts.

2.6 Note that, because this procurement is funded by a Department of Education grant, Mann Library is eligible for all pertinent GSA discounts.

2.7 Also note that, because Mann Library is a unit of the statutory colleges at Cornell University, this procurement process is subject to all pertinent regulations and purchasing restrictions of the SUNY system and the State of New York, as well as standard Cornell University purchasing procedures. Administrative aspects of this procurement process will be managed by Cornell Purchasing (contact: Don Dill, tel. (607)-255-7409). Technical aspects of the procurement process, and any and all systems evaluation, will be conducted by Mann Library with the consultation of Cornell Information Technologies and CISER, the Cornell data archive (contact: Howard Curtis, tel. (607)-255-9570).

3. Description of evaluation and procurement process:

3.1 Vendors shall have 30 days for the preparation of proposals. The due date is specified in the cover letter accompanying this document. Questions concerning the due date should be directed to Don Dill at Cornell Purchasing.

3.2 Mann Library will host an informational session on this request for proposal, to take place approximately half-way through the response period. At this session, which vendors are free to attend or not to attend, at their choice, library and technical staff will answer vendor questions. The time and place of this meeting will be announced in the cover letter accompanying this document. The library will not provide a transcript of this meeting to vendors who choose not to attend.

3.3 Mann Library will entertain questions in writing concerning this request for proposal for the first 14 days of the response period. We will provide written answers to these questions, together with the texts of the original questions, to all vendors. Vendors should direct such questions to: Howard Curtis, Information Technology Section, Mann Library, Cornell University, Ithaca, New York 14853.

3.4 All bids will be considered on both technical and financial grounds in each of the areas of concern identified in this bid solicitation. In conducting our proposal evaluation, we will consider the following areas, at minimum:

3.4.1 System performance specifications
3.4.2 Price
3.4.3 Compatibility with the existing Cornell computing environment (by which is meant an ability to integrate a vendor's offering into the existing computing environment)
3.4.4 Networking capabilities
3.4.5 Cost of necessary software (i.e., if the library needs to license third-party software to achieve its goals, in addition to vendor-supplied software, how much will the entire package cost?)
3.4.6 Expandability of system
3.4.7 Vendor services and maintenance options
3.4.8 Vendor history

3.5 The evaluation of proposals will be undertaken by an evaluation team consisting of library staff members and technical staff from Cornell Information Technologies and CISER.

3.6 The library will invite vendors deemed competitive in an initial review of the written proposals to make a presentation to the Cornell evaluation team concerning their proposals. These presentations will take place at Mann Library, at a time to be scheduled with the individual respondents.

3.7 Mann Library and Cornell University reserve the right to reject any and all vendor proposals, at their sole discretion, on technical or financial grounds, or on other criteria for evaluation, including but not limited to those listed above.

3.8 Mann Library will conduct this evaluation process in a fair and impartial manner, based on its understanding of system requirements as outlined here. Mann Library will make a public announcement of its choice of a computer system at the end of the evaluation process. The library, however, is under no obligation to justify its choice of a system to any participating vendor.

4. Systems Requirements

4.1 Central Machine

4.1.1 The library requires a central processing unit capable of supporting 20 simultaneous sessions. The central unit may be either a single or multiple-processor configuration.

4.1.2 The central processing unit should include at least 16 Megabytes of RAM. Please specify amount of RAM in the configuration proposed.

4.1.3 The library is interested in the degree to which the proposed processing unit can be expanded or upgraded. Specify expansion and upgrade options and current prices. Your account should include addition of processors, RAM modules, and other significant options.

4.2 Secondary Storage

4.2.1 The library requires a minimum of 500 megabytes of magnetic disk storage. Please specify magnetic disk prices in increments from approximately 500 megabytes to 2 gigabytes. Specify make of drive, capacity, access time, price, and other pertinent specifications of drive(s) proposed. Also specify the maximum amount of magnetic storage supported by the proposed processor architecture.

4.2.2 The library is interested in the prospect of mounting full-text files on optical media. Please quote an optical drive subsystem, if available (this optical drive subsystem can be a product of a third-party). Specify make of drive, capacity, access time, price, and other pertinent specifications of drive(s) proposed. Specify whether a juke-box system is available for the the drive(s) proposed.

4.3 Offline Storage

4.3.1 The library requires the capability to load data files from magnetic tape. This may involve the purchase of a tape drive, or the use of a CIT tape drive through appropriate network connections.

4.3.2 Please quote a magnetic tape drive with the following characteristics: capable of accepting 9-track magnetic tape on reels; 1600 or 6250 bpi. Specify make of drive, price, and other pertinent specifications of drive proposed. If a method of employing a tape drive currently operated by CIT is known to you, please explain.

4.4 Networking Considerations: It remains unclear whether the computer that Mann Library acquires will be located in the library building or in some other building on campus (for example, under CIT systems management in the CCC building). For the purpose of preparing a bid, assume a location in Mann Library. Then specify what additional equipment would be required if the machine should be located in another building, and what facilities Cornell would have to provide in order to make an alternative site possible.

Mann Library currently has in place a radial configuration of 11 lengths of "Thinwire" Ethernet cable running from the library's central communications hub to staff work areas. These cable runs are available for use in support of system communications.

The library also has in place a radial (star) network of twisted pair runs which is in use to support terminal communications with the central CIT computers over the Sytek (broadband) network.

If you require additional information on the current networking situation within Mann Library or on the Cornell campus, please contact Howard Curtis.

Networking (Information server and library workstations)

4.4.1 The library seeks a highly integrated network communication package for 1 IBM-AT and two AT-compatible computers located in two rooms of the library (both with access to "Thinwire" Ethernet as outlined above). Please quote hardware and software components required.

4.4.2 The library seeks one communications connection to an Appletalk network. This network is currently limited to Macintosh hardware, but in the future will likely include both Macintosh and PC-compatible machines running under TOPS or Novell Netware software.

4.4.3 The library seeks one communications connection to support terminal sessions at up to 8 PCs running terminal emulation software.

Networking (Information server and campus network)

4.4.4 Mann Library desires that its computer support a peer-to-peer connection to the CIT "mainframe establishment," consisting of IBM mainframes running VM/XA and DEC machines running VMS, through the campus telecommunication network. Of particular concern is the ability to transfer data at high speed from CIT tape drives to the magnetic disk storage of the Mann Library computer.

4.4.5 Please specify, as closely as is known, the hardware and software components required to achieve this connection, with price information.

4.5 Operating System Software

4.5.1 The library requires an operating system software license for up to 10 simultaneous users. The operating system should come with a standard complement of utilities for system administration and network control. Mann Library is interested in the following operating systems: (1) Unix, (2) VM/CMS, and (3) VMS. Please quote an operating system for your proposed system with the following information: operating system, current version, initial licensing charges, annual maintenance fee, charges for processor upgrades. [Note: If you offer two of the library's desired operating systems, please quote both].

4.5.2 The library seeks a network management package that would allow us to monitor and troubleshoot the networking facilities described in the "Networking (Information server and library workstations)" section above.

4.5.3 The library seeks to support DOS and Macintosh file services on the minicomputer. This facility would allow library staff members to store DOS and Macintosh program and data files on the minicomputer's magnetic disk and to transfer them to other staff members. Please quote the cost of software necessary to achieve this, if available.

4.6 Applications Software

4.6.1 The library seeks a database management system that would provide the core capabilities of the interactive data extraction system described above and in Attachment 1.

4.6.2 The library seeks a programming environment to support work in the C language, to include editing, linking, compiling, and debugging capabilities.

[Note: Although vendors are not responsible for quoting individual pieces of applications software, they should be aware, that the library will be highly sensitive to the availability and cost of high-quality database and language products. Vendors are encouraged to quote prices for their own software packages or for third-party products that they believe would facilitate the system development work described in the two attachments to this proposal. Vendors with agreements in place or pending which provide access to large "libraries" of applications and/or systems software at reasonable cost should state the terms under which Mann Library could license software under this plan.]

4.7 Warranty and Maintenance

4.7.1 Please state the terms of the warranty coverage that applies to the various components of the system proposed. Explain where and how service under warranty is carried out; explain who performs necessary service work.

4.7.2 Please explain maintenance options on the system proposed with a clear account of annual charges for each option or level.

4.8 Installation

4.8.1 It is not presently clear whether the computer system that Mann Library procures will be located in the CCC (Cornell Information Technologies' Computing and Communications Center) or in the Mann Library building. Vendors should thus assume the possibility that hardware will operate in a non-computer room environment. Please include with your proposal specifications for the necessary operating environment for your system and data on the environmental characteristics of your system. This information should include temperature ranges, electrical requirements, footprint, BTU ratings, etc.

4.8.2 Mann Library seeks to have in place a functional system with active network connections to library staff areas as outlined above by July 1, 1989. Please specify lead time to delivery and time required for key installation procedures. Specify infrastructure requirements that the library should prepare in advance (electricity, cable installation for networking support, air conditioning, site preparation, etc.).

4.9 Leasing Possibility

4.9.1 All interested vendors should, at a minimum, quote price-to-purchase figures in their bids. As the Department of Education grant in question does permit leasing arrangements, however, the library is also open to this possibility. Vendors are encouraged to quote leasing rates and terms, should they feel that leasing would represent an attractive possiblity in this case.

4.10 Ongoing Cost

4.10.1 In addition to your bid on the system components listed here, please provide your best estimate of the 3-year cost of ownership of the overall system you are proposing. This estimate should include hardware maintenance and software licensing fees. If a bid specifies figures for both price-to-purchase and leasing arrangements, the vendor should provide 3-year cost of ownership estimates for both arrangements.

Mann Library--Information Server Procurement
Summary of Requirements

Section	Item
Central Machine	
4.1.1	Central processing unit, capable of supporting up to 20 simultaneous users
4.1.2	RAM memory module of at least 16 megabytes
4.1.3	Expansion options for central processing unit, RAM, etc.
Secondary Storage	
4.2.1	Magnetic DASD in base amount of approx. 500 megabytes
4.2.1	Specification of cost of increments of magnetic DASD to 2 gigabytes
4.2.1	Maximum amount of secondary magnetic storage supported
4.2.2	Optical disk subsystem
4.2.2	Availability of optical juke box configuration
4.3.2	Magnetic tape drive
4.3.2	Ability to use tape drive currently available in Cornell's CIT facilities
Networking Capabilities	
4.4.1	Integrated networking package for 1 IBM PC-AT and two PC-AT compatible microcomputers
4.4.2	Connection to Appletalk network
4.4.3	Communications connection to support terminal-emulation sessions at 8 IBM PC-compatible computers
4.4.4	Communications connection to other Cornell central computers (connection to the campus backbone)
Software	
4.5.1	Operating system software (Note possibility of quoting two operating systems)

4.5.2	Network management software
4.5.3	Software to provide DOS and Macintosh file services to networked microcomputers
4.6.1	Database management software
4.6.2	Programming environment support software

Warranty and maintenance

4.7.1	Warranty coverage
4.7.2	Maintenance provisions

Specifications and installation

4.8.1	Operational specifications
4.8.2	Lead time to delivery; time requirements for installation
4.8.2	Infrastructure requirements
2.2	Installation charges and other charges for vendor-supplied, one-time technical work
2.3	Installation services and other technical, start-up activities that Cornell University personnel must undertake

Leasimg option and 3-year cost

4.9.1	Leasing possibility
4.10.1	3-year cost of ownership

Appendix B: Request for Proposal — Software

Section I

Background Information - Interface Development

The files to be incorporated into the system will be acting as data archives made available for data extraction and manipulation. Each file will constitute a database with 1+ tables. Currently, the largest database will be comprised of up to 12 tables. Therefore, 4 essential system functions are necessary:

1. The ability to select the fields to base the search on and enter the necessary search criteria
2. The ability to select the fields desired for display or file extraction
3. Allow as much logical grouping control as possible with the use of AND's, OR's, ()'s, etc ...
4. As part of the application, and as easily as possible, extract and export the resulting data (in columnar form). The more file formats available, the better, with a minimum being an ascii file with delimiter of choice so that the data can be imported to various software packages. Multi-table extractions to flat file format would need to be accommodated.

The user interface must be designed to accommodate large numbers of fields for query creation (as in #1. #2, and #3 above). We are attempting to create an environment that allows picking fields of interest from lists of options in the most organized manner by possibly moving in and out of menu items to make selections and build the SQL query. Unlike a standard reporting module where data access points can be easily identified (ie. account #, date range, product #, etc ...), the necessary access points will vary from user to user and session to session.

Question

Please be prepared to explain and demonstrate how an interface of this type might be achieved using your product tools. The idea of dynamic SQL incorporating "placeholders" (to be substituted for during data entry of some sort) for this type of result may not have the amount of user control & flexibility we wish to achieve, due to the limitations in the number of search conditions and output field placeholders that may be incorporated. An SQL statement that can be built from scratch, internally, using the data selected and entered (string concatenation?), seems to be what we would like to achieve. We have some doubts about the applicability of Query-By-Example or Query-By-Forms to our application, especially if a static data entry-type screen must be incorporated which may result in excessive <return>'s through unused fields. Please explain where it would be appropriate, if you think it would be.

In your discussion, we would prefer that the distinction be made between what can be achieved within your 4GL environment and what may require the use of a 3GL to meet these needs. Please note the merits of each. Actual code that demonstrates these elements would be most desireable.

Can important field information such as statistical sample weights, flags and footnotes (textual information necessary to qualify individual data values, provide explanation for missing data or outlying data, etc ...) be programmed to be automatically taken into consideration when a field requiring this validating information is selected or queried? This authoritative information will not necessarily be apparent to the user, but is critical data. Also, this information may be stored in a related table, so should most likely be queried automatically when a field in question is selected.

Will, or can an exported file contain the above descriptive field information and column (field) headings, or must that information be exported separately and left to the user to put together for use with other software?

Background Information - Indexing & Database Access considerations

Another issue is database access time. Obviously, an application that produces very defined output and functions (ie. specialized reporting), can be optimized to a large degree by the developer by strategically constructing and hardcoding the various components of a search statement into a program. Our application is attempting to cater to a variety of ad-hoc needs. We wish to allow basically free-format search statement generation by picking, choosing and typing search criteria data. With this approach, hardcoding seems to be out of the question, but dynamic statement generation is not. We are trying to get a feel for the consequences of this approach on access time and indexing decisions.

Questions

Please be prepared to discuss the degree to which your Query Optimizer takes into consideration the order of the syntax elements as they are interpreted, or whether solely what is being requested, not how it is being requested is of most importance when choosing the ultimate search path.

We would appreciate a discussion on your product's query optimization approach to determining the best search path. Please use examples where necessary.

If space permits, and the overhead of numerous indexes is not a factor, would it be advantageous to index all particularly heavily-used search fields if such a "list of heavily accessed fields" could be determined, even if this resulted in 50+ indexes for a particular database?

Please explain how the following example searches on fields A,B,C and D would be handled. Please note the placement of logical and relational operators.

- If all fields were indexed, how would the indexes be utilized to achieve the queries?
- If fields B & D are historically not heavily searched fields, and only fields A & C are indexed, how would that effect the searching strategy?

```
SELECT ... WHERE (A = ? and B = ?) OR (C = ? and D > ?)
SELECT ... WHERE A = ? and ( (B > ? and B < ?) or C = ? or D = ?)
SELECT ... WHERE (A = ? or B = ?) and (C = ? and D = ?)
```

Section III

Background Information - Null Values

A majority of the files to be incorporated contain statistical data, and quite often there are occurrences of missing data (either the data was not available or could not be published due to confidentiality, etc ...). The treatment of this data is very important and may vary. The preference concerning how records containing null value data should be handled, especially if the field(s) in question is/are part of a search condition, should be left very much up to the individual user session. Some users may wish to inspect data records that do not meet a specified search condition due to a missing data value, others may not.

For example, some users may wish to overlook records with missing data in search fields because they are only interested in records for which data exists. Other users, on the other hand, may want to see records with missing data in search fields to further investigate the reason for the missing data. In this case, the user would want data records with missing data in search fields to be included with all other matching records for a particular search.

Questions - Null Values

In general,

1. How does your product support Null values (missing data) vs. "0" and "spaces" in Numeric and Character fields, respectively? Please be specific. Will a record containing a null value in an index field be included in the index and be accessible as part of a search (excluding the primary key)?
2. Depending on the result of #1, can the choice to include/exclude null data during a search (or computation) be controlled to suit an individual session? Can it be done at the application level as some sort of environmental parameter?
3. What is the impact of missing data on statistical calculations?

Using the example data below, please explain how the following statements would be handled and what the results would be according to your product's treatment of Null values. Note: The data is fictitious, and the # of occurrences of missing data is accentuated.

Commodity Code	Year	State	Yield	Produc- tion	Area Planted	Harvested
1	82	PA	39	2106	55	54
1	82	NY	34	30940	.	910
1	82	NJ	67	55610	1000	830
1	83	PA	.	.	.	.
1	83	NY	46	64400	1550	1400
1	83	NJ	38	.	2200	1850
2	82	PA	34	20400	720	600
2	82	NY	.	7500	430	250
2	82	NJ	35	2625	100	75
2	83	PA	.	.	.	.
2	83	NY	30	7500	430	250
2	83	NJ	41	5371	145	131

1. # Records selected for (YIELD >= 38)
2. Average Yield for (YIELD >= 38)
3. # Records selected for (YIELD >=30 and YIELD <= 35)
4. # Records selected and Average Production for
   (YIELD >=30 and YIELD <= 35) and (PLANTED > 400)
5. # Records selected and Average Production for
   (YIELD >=30 and YIELD <= 35) or (PLANTED > 400)
6. If any records containing missing data were exported to a text file, how would the missing data be represented? In any other supported file formats?

Section IV

Questions to Vendors

1. Please be prepared to discuss any facilities for TCP/IP networking, and network communications in general.
2. Does your product support HP-UX long file names?
3. We anticipate that a large number of our users will initially access the HP server via terminal emulation. Please explain any possible issues that arise when a terminal accesses your database product at low baud rates.
4. Are there any extra space requirements for reformatting or reindexing data?
5. What special data preparations must be performed to load a non-delimited, ascii file into a database structure? What if the data to be loaded is to be loaded into a multi-table database?
6. As a future possibility, we may wish to explore the idea of relating two or more data files (database). For example, the National Resources Inventory provides various soil sample data for a particular survey year. The Toxic Release Inventory data provides chemical release information for various facilities for a particular reporting year. Both of these files provide FIPS State & County data. Would initially defining these two files to be independent databases require substantial restructuring and attention if we were to later wish to provide TRI and NRI data to a user-specified area of geography? Or can queries be defined to extend across 2+ databases?
7. Please discuss any run-time user environment issues.
8. The application will allow exporting data subsets for further analysis within statistical packages residing on the user workstation. Please discuss any file transfer issues at the application level vs. the operating system level for:
    - accessing the server as a dumb terminal
    - accessing the server from a front-end application
9. In a multi-table database, does the total # of tables containing search statement elements during a query have an impact on eventual access time vs. a one-table database?
10. Provide an example of a "complex" SQL statement that your product handles particularly well.
11. If your product supports a workstation front-end building environment, can the front-end application communicate with the server through an asynchronous line with the so-called SLIP protocol?
12. When an applications developer has a technical problem, what is the procedure to get assistance?
13. Please estimate software maintenance and vendor support costs for the first 3 years of operation.

SECTION V

BACKGROUND INFORMATION - CUSTOM HELP

We are currently interested in developing 4 types of custom help to facilitate the use of this application at all levels of the application. Please respond to the various types and how they might be implemented.

1. Your basic "opening screen", not soliciting any action. To be viewed, then removed with a "Press any key to continue" feature.

> National Resources Inventory data files for 1988 have not been received, when they arrive they will be loaded into the database for searching.
> Press any key to continue ...

2. Further data definitions, descriptions or valid entries from the documentation provided by the producers of the database. This form of help we would like to make available at 2 levels:
   a. Form-based help with a hierarchical capability for various levels of descriptive help within a particular topic. For example, select Help for a particular category of fields from a menu of categories, that would then allow viewing descriptive Help text for the various individual fields within that category.
   b. Field-based help that would be accessed at the field level (on a data entry screen of some sort, for example) giving descriptive information as well as valid data entry codes and/or descriptions. How much of this information could be extracted from the Data Dictionary, how much might have to be hardcoded somewhere, and how much could be stored in a table external to the data, accessed from there and incorporated into the help screen? Raw data files are quite often accompanied by "code book" data containing descriptions of various field contents, footnotes,flags, etc ... This information should be available for view to assist in data entry (if the individual does not know the allowable codes or what they represent, for example) or simply as a description of the variable.

> **Land Disposal Code**
> A code corresponding to the type of land disposal used for the toxic chemical at the facility:
> D02 - Landfill
> D03 - Land Treatment/Application/Farming
> D05 - Surface Impoundment
> D99 - Other

3. For fields with a large number of possible entries (State abbreviations, FIPS State & County Codes, Chemical codes, etc ...), a screen similar to that in # 2 that would also allow scrolling and data entry selection from within the box would provide very useful. The screen should hold a reasonable # of elements and allow scrolling for more.

4. The ability to estimate the access time of a search and make the user aware of the estimated time required with the option to abort the search request is very desireable. Aborting a search at any time should be made possible whether or not the ability to estimate search time is possible.

in MicroCenter

DBMS issues yet to be addressed:

1. 4GL ability to generate a series of menus and sub-menus accessing several QBE forms (to categorize and manage the large volume of fields available for access) to jointly CONSTRUCT the SQL search statement.

   Move in and out of various field category data entry screens, and press button to process data meeting any/all entries made to all field category screens./

2. 4GL ability to dynamically incorporate field EXPORT selection as part of the above CONSTRUCT process.
   ie. SELECT __________ WHERE __________

3. 4GL abillity to control AND & OR operators BETWEEN fields as well as within fields
   ie. (field1 = a OR field2 = b) AND field3 = c

4. Can a 4GL application be designed to program cursor control (sequence of fields entered during data entry) and general screen elements in the form of a QBE screen or something similar if the QBE feature is not flexible enough for our needs?

5. Are most people able to find the query flexibility they need in a read-only, read-intensive environment?

6. For an ad-hoc approach, is it recommended to force particular index use by controlling allowable primary field access points that coincide with existing indexes when dealing with large data files with lots of fields?

   If so, can sub-queries then be generated on the temporary file created by a primary search statement? Would this occur while still in the same application? How would this new temporary dataset be accessed?

   Distinction between sub-queries and joins and there impact, if any.

7. Can dynamic SQL be constructed by concetenation based on user entries, or must it incorporate placeholders that restrict the # of arguements?

8. Can important field information such as sample weights, flags and footnotes be programmed to be automatically taken into consideration when a field requiring this validating information is selected or queried? This authoritative information will not necessarily be apparent to the user, but is critical data. Also, this information may be stored in a related table, so should most likely be queried automatically when a field in question is selected.

If NO to any of the above items, can it be incorporated as a 3GL extension to the 4GL application or does it entail an entire 3GL interface. If so, what is the complexity of creating such a module.

1. a. Once a query has been performed, what are the exact methods that can be tied into the application to export this data set?

   b. If the ability to select output fields to be exported exists (except the assumed default of "all"), can this information also be easily tied into the application?

   b

   c. will/can an exported file contain descriptive field information or any associated "flags" or "footnotes", or must that information be exported separately and left to the user to put together?

   d. can FTP be performed from within application?

2. Can the 4GL application be developed to begin with an opening menu allowing the user to select which database they wish to search, then move right into the code for that particular database? As opposed to having to develop individually-created applications that would require the user to exit one, then run another application to search another database.

   c

3. Each file we are loading will constitute a database with 1+ tables. If applications cannot be built to handle more than one defined database at a time, and 2 or more independent files have a common element and have some data relationship or validity together (ie. FIPS codes, NRI & TRI), could they be easily or sensibly combined to become a database with tables distinguishing them and allowing individual searching, but also allow the opportunity to extract NRI & TRI data for "New York State"?

   Is this not necessary for software that claims to "not require programs to be saved with a database?"

4. Please provide more detail on the products query optimization methods and their impact on large file access in the event of a "bad" search. Also, comment on the impact of a multi-table database in a "good" and "bad" search.

   Do the total # of tables containing search criteria elements have an impact on eventual access time?

   Request complex SQL statements that the products work well on.

5. Any inherent safegaurds against searches that could potentially request entirely too much data to be handled? Is this even a concern?

6. Can a Report Writer program be passed a list of selected fields and drop them into placeholder columns? If so, can stats be performed if numeric? Is there the ability to save "if numeric, average ...", or better yet allow user-input formula's consisting of field names + operators, and have the Report Writer dump into columns?

7. Can a friendly, helpful SQL environment be developed that has the feel of an application (help, etc ...), but results in the user

constructing and entering an SQL statement if they are familiar with both the database table structures and SQL?

Appendix C: Sample Data Output, comma delimited

Sample Data Extraction from the National Resources Inventory Partial record output, comma delimited

```
24,003,0126054,2,0010,"149A",02,06,00,06,000008,18,1,2,
24,003,0126054,3,0010,"149A",02,06,00,06,000009,18,1,2,
24,003,0126061,2,0010,"149A",02,06,00,06,000009,10,1,2,
24,003,0126061,3,0010,"149A",02,06,00,06,000008,10,1,2,
24,005,0126131,3,0210,"149A",02,06,00,03,000006,12,2,2,
24,005,0126179,1,0210,"149A",02,06,00,03,000009,12,1,2,
24,009,0126256,3,0160,"149A",02,06,00,06,000006,16,1,2,
24,009,0126267,2,0160,"149A",02,06,00,06,000006,17,1,2,
24,009,0126272,3,0160,"149A",02,06,00,06,000007,16,1,2,
24,009,0126291,2,0160,"149A",02,06,00,06,000006,12,1,5,
24,009,0126291,3,0160,"149A",02,06,00,06,000006,12,1,5,
24,009,0126308,1,0160,"149A",02,06,00,06,000006,17,0,2,
24,011,0126323,3,0074,"153C",02,06,00,05,000006,12,1,5,
24,015,0126583,1,0093,"149A",02,06,00,02,000008,12,1,5,
24,015,0126583,2,0093,"149A",02,06,00,02,000008,12,1,5,
24,015,0126583,3,0093,"149A",02,06,00,02,000009,12,1,5,
24,017,0126647,2,0130,"149A",02,07,00,11,000009,12,1,5,
24,017,0126647,3,0130,"149A",02,07,00,11,000009,12,1,5,
24,017,0126648,1,0130,"149A",02,07,00,11,000008,12,1,5,
24,017,0126648,3,0130,"149A",02,07,00,11,000009,12,1,5,
24,017,0126652,1,0130,"149A",02,07,00,11,000009,12,1,5,
24,017,0126661,1,0130,"149A",02,07,00,11,000008,12,3,5,
24,017,0126696,1,0130,"149A",02,07,00,11,000009,12,1,2,
24,017,0126697,2,0130,"149A",02,07,00,11,000009,12,1,5,
24,017,0126697,3,0130,"149A",02,07,00,11,000008,12,1,5,
24,019,0126737,2,0032,"153B",02,06,00,01,000010,17,1,2,
24,019,0126738,3,0032,"153B",02,06,00,05,000010,16,1,2,
24,019,0126739,1,0032,"153B",02,06,00,07,000010,16,1,2,
24,019,0126739,2,0032,"153B",02,06,00,07,000010,16,1,2,
24,019,0126740,1,0032,"153B",02,06,00,07,000010,17,1,2,
24,019,0126740,2,0032,"153B",02,06,00,07,000010,17,1,2,
24,019,0126740,3,0032,"153B",02,06,00,07,000011,17,1,2,
24,019,0126741,1,0032,"153B",02,06,00,07,000010,16,1,2,
24,019,0126741,2,0032,"153B",02,06,00,07,000010,16,1,2,
24,019,0126741,3,0032,"153B",02,06,00,07,000010,16,1,2
24,019,0126746,1,0032,"153B",02,06,00,08,000010,17,1,2,
```

Appendix D — Full Codebook

This is the file output with the .inf extension when the full codebook is requested from the Run menu. Yellow highlighted sections are what would be output for the search if no codebook had been requested and only wetland related fields had been selected for output.

```
+---------------+
| ** WARNING ** |
+---------------+
```

Draft Disclaimer (revised)

The 1982 NRI was designed primarily to obtain natural resource data usable for analysis of non-Federal land at a substate (multi-county) level. Estimates based on these data can be developed for such substate entities as Major Land Resource Areas (MLRAs), SCS Administrative Areas, sub-river basins, and Water Resources Council aggregated subareas (ASAs). The sample was selected specifically for use in analysis at the MLRA within state level.

Even though the 1982 NRI was designed for MLRA level analysis, the data can be used to easily compute numerous estimates -- such as, for example, acreage by land use category -- for individual counties, hydrologic units, and similar small-sized areas. Many estimates at this level are not considered reliable enough to recommend their use for decision making. The following example illustrates how reliability decreases as the unit of analysis becomes smaller:

Reliability of Pastureland Acreage Estimates

Region	Estimated Acres	95% Confidence Interval
U.S.	133,310,000	+/- 0.8%
Iowa	4,536,000	4%
MLRA 107	1,770,000	6%
Adair County, Iowa	76,000	29%

It is also important to realize that each item being estimated has a different level of precision (or reliability). Characteristics that are common and spread fairly uniformly over the region of interest exhibit smaller relative variation than those characteristics that are rare and unevenly distributed.

(Adapted from National Resources Inventory: A Guide for Users of 1982 NRI Data Files, 1984.)

```
+---------------+
|  Supporting   |
| Documentation |
+---------------+
```

For the file:

Goebel, J. Jeffrey, and Richard K. Dorsch. (October 1984, Revised

July 1986) National Resources Inventory. A guide for users of 1982 NRI data files.
(Available from the Soil Conservation Service, or from Interlibrary Services, Mann Library, Cornell University, Ithaca, NY 14853-4301)

For Field #5: Major Land Resource Area
United States Department of Agriculture. Soil Conservation Service. (December 1981) Land Resource Regions and Major Land Resource Areas of the United States. Agriculture Handbook 296. Washington, D.C.: Government Printing Office.

For Field #6: Hydrologic Unit
United States Geological Survey. (1982) Codes for the Identification of Hydrologic Units in the United States and the Caribbean Outlying Areas. A U.S. Geological Survey Data Standard Geological Survey Circular 878-A. Alexandria, VA: U.S. Geological Survey.

For Field #9a: Land capability class
United States Department of Agriculture. Soil Conservation Service. (1961) Land-capability classification. Agriculture Handbook 210. Washington, D.C.: Government Printing Office.

For Field #48: Wetland type
Shaw, Samuel P., and C. Gordon Fredine. Wetlands of the United States. Their extent and their value to waterfowl and other wildlife. (1956) Circular 39. Fish and Wildlife Service, United States Department of the Interior. Washington, D.C.: Government Printing Office.

```
+-------------------------+
| For Further Information |
+-------------------------+
```

For further information about the content of the files please contact the Resources Inventory Division, Soil Conservation Service, U.S. Department of Agriculture, 12th & Independence Ave., S.W., Washington, D.C. 20013.

For further information about the loading of the file within the Aqueduct system, or information about Aqueduct in general, please contact the Computer Files Service, Mann Library, Ithaca, NY 14853-4301, (607) 255-2199.

```
+---------------------------------------------------------------------+
|                    Fields included in your dataset                  |
+---------------------------------------------------------------------+

+-----------------+
| Fields in order |
|    of output    |
+-----------------+

 #  Field Name                                                    Start End

 1  FIPS State code                                                  1    2
 1  FIPS County code                                                 3    5
 2  PSU #                                                            6   12
 3  Point #                                                         13   13
 4  SCS Location code                                               14   17
 5  MLRA                                                            18   21
 6  Hydrologic Unit - Region                                        22   23
 6  Hydrologic Unit - Sub-Region                                    24   25
 6  Hydrologic Unit - Accounting Unit                               26   27
 6  Hydrologic Unit - Cataloging Unit                               28   29
 7  Expansion Factor                                                30   35
 8  Ownership of Land                                               36   36
 9  Land Capability Class (9a)                                      37   37
 9  Land Capability Sub-Class (9b)                                  38   38
10  T-Factor                                                        39   39
11  Prime farmland?                                                 40   40
12  Degree of erosion                                               41   41
13  Nonarable (due to past erosion)?                                42   42
14  NA                                                              43   43
15  Saline and/or alkali soil (special management need?)            44   44
16  Type of irrigation                                              45   45
17  Water source for irrigation                                     46   46
18  Water provision; irrigation at least 1/2 of water requirements? 47   47
19  Flood prone area?                                               48   48
20  Cover/use, general                                              49   49
21  Cover/use, major                                                50   50
22  Cover/use, specific                                             51   53
23  Use of the land or water                                        54   55
24  Cropping history 1981                                           56   58
25  Cropping history 1980                                           59   61
26  Cropping history 1979                                           62   64
27  Double-cropping used?                                           65   65
28  Conservation practise 1                                         66   68
29  Conservation practise 2                                         69   71
30  Conservation practise 3                                         72   74
```

31	Conservation treatment needed	75	76
32	K-Factor	77	78
33	R-Factor	79	81
34	C-Factor	82	87
35	P-Factor	88	90
36	Slope length for Universal Soil Loss Equation (USLE)	91	94
37	Slope percent for Universal Soil Loss Equation (USLE)	95	98
38	Universal Soil Loss Equation (USLE) flag	99	104
39	Universal Soil Loss Equation (USLE)	105	105
40	Universal Soil Loss Equation (USLE) tons	106	114
41	Wind erosion	115	120
42	Wind erosion tons	121	129
43	Dominant soil and water problem	130	131
44	Secondary soil and water problem	132	133
45	Dominant non-soil or water problem	134	135
46	Type of effort necessary for conversion to cropland	136	137
47	Conversion potential	138	139
48	Type of wetland	140	141
49	Kind of wetland vegetation	142	142
50	Kind of wetland system	143	143
51	Riparian area kind	144	144
52	Riparian vegetation kind	145	145
53	Riparian vegetation; width of strip	146	146
54	Distance to cropland	147	150
55	Distance to forest land	151	154
56	Distance to grassland	155	158
57	Distance to water	159	162
58	Distance to wetlands	163	166
59	Distance to built-up land	167	170
60	Winter cover	171	171
61	Winter cover height	172	173
62	Residue remains upright over winter?	174	174
63	Pastureland condition rating	175	175
64	Woody canopy cover, for pastureland	176	176
65	Rangeland condition rating (percent climax vegetation)	177	177
66	Woody canopy cover, for rangeland	178	178
67	Rangeland condition trend	179	179
68	Grazing level, for rangeland	180	180
69	Forest type, general category	181	181
70	Forest type, specific category	182	184
71	Canopy cover of the trees, for forest land	185	185
72	Basal area/stem count	186	188
73	Diameter at Breast Height (DBH)	189	191
74	Forest understory composition	192	192
75	Understory forage value	193	193
76	Soils-5	194	245

```
+----------------------------+
| Geography Search Conditions |
+----------------------------+

FIELD #: 1 = FIPS State/County

24       State - MARYLAND

  001    ALLEGANY
  003    ANNE ARUNDEL
  005    BALTIMORE
  009    CALVERT
  011    CAROLINE
  013    CARROLL
  015    CECIL
  017    CHARLES
  019    DORCHESTER
  021    FREDERICK
  023    GARRETT
  025    HARFORD
  027    HOWARD
  029    KENT
  031    MONTGOMERY
  033    PRINCE GEORGES
  035    QUEEN ANNES
  037    ST MARYS
  039    SOMERSET
  041    TALBOT
  043    WASHINGTON
  045    WICOMICO
  047    WORCESTER


+-------------------------+
| Other Search Conditions |
+-------------------------+

FIELD #: 48 = Type of wetland

  FILTER: Yes.  Filter conditions marked with an asterisk (*).

    Code     Label
    ----     -----
      00   None
      01   Seasonally flooded basins or flats
           Few inches in upland; few feet along rivers
```

02	Inland fresh meadows Few inches after heavy rains
03	Inland shallow fresh marshes Up to 6 inches
04	Inland deep fresh marshes Up to 3 feet
05	Inland open fresh water Up to 10 feet; marshy border may be present
06	Shrub swamps Up to 6 inches
07	Wooded swamps Up to 1 foot
08	Bogs Shallow ponds may be present
*09	Inland saline flats Few inches after heavy rain
*10	Inland saline marshes Up to 3 feet
*11	Inland open saline water Up to 10 feet; marshy border
*12	Coastal shallow fresh marshes Up to 6 inches at high tide
*13	Coastal deep fresh marshes Up to 3 feet at high tide
*14	Coastal open fresh water Up to 10 feet; marshy border often present
*15	Coastal salt flats May have few inches at high tide
*16	Coastal salt meadows May have few inches at high tide
*17	Irregularly flooded salt marshes Few inches at high tide
*18	Regularly flooded salt marshes Up to 1 foot at high tide
*19	Sounds and bays Up to 10 feet at high tide
*20	Mangrove swamps Up to 3 feet

```
+--------------------------------------------------------------------------+
|                      Codebook for all output fields                      |
+--------------------------------------------------------------------------+

FIELD #:  1 = FIPS code
  State and county Federal Information Processing Standard (FIPS) code

FIELD #:  2 = PSU #
  Primary Sampling unit (PUS) number

FIELD #:  3 = Point #
  Specific point within the PSU where the inventory data are collected--
  a data record is not included in the NRI file for points that are
  federal, urban and built-up, transportation facilities, water bodies, or
  streams.

FIELD #:  4 = Location code
  SCS location code

FIELD #:  5 = Major Land Resource Area (MLRA)
  Major Land Resource Area (MLRA) as per USDA Agricultural Handbook 296
  (Dec. 1981)

FIELD #:  6 = Hydrologic Unit
  Water Resource Council Hydrologic Unit -- 8-digit cataloging unit number,
  as per U.S. Geological Survey Circular 878-A (1982)

FIELD #:  7 = Expansion Factor
  Number of acres the sample point represents (in 100s) -- it takes into
  account the sampling procedure and the state's census acres; it is the
  figure to use when constructing acreage estimates (for categories in which
  the point falls)

FIELD #:  8 = Ownership of land
   Code     Label
     0   Not applicable
     1   Private
     2   Municipal
     3   County or Parish
     4   State
     5   Federal (not in file)
     6   Indian tribal and individual trust lands
```

FIELD #: 9a = Land capability class

Soils suitable rating for agriculture, between 1 and 8, as per Agriculture Handbook 210

Code	Label
1	Soils with few limitations that restrict their use
2	Some limits--less plant choice or mod. conserv. practice req
3	Severe limits--less plant choice +/or special conserv. req.
4	Very severe limits on plant choice/special conserv. mgmt.
5	Cultivation impractical without major reclamation
6	Very severe limits--unsuited to cultivation
7	Extreme limits/restrict to wood,wildlife,spec.mgmt. pasture
8	Suited only to wildlife, recreation, water supply, esthetics

FIELD #: 9b = Land capability class; soil limits

Chief limitation of the soil (except when class 1)

Code	Label
C	Climate
E	Erosion
S	Shallow, drought, or stony
W	Water

FIELD #: 10 = T-Factor

Soil loss tolerance factor -- indicates acceptable level of annual soil loss, between 1 and 5 tons per acre per year

UNITS:

MISC: tons

FIELD #: 11 = Prime farmland?

Meets prime farmland criteria?

Code	Label
1	yes
2	no

FIELD #: 12 = Degree of erosion

Code	Label
1	None or slight
2	Moderate
3	Severe

FIELD #: 13 = Nonarable (due to past erosion)?

Code	Label
1	yes
2	no

FIELD #: 14 = NA

FIELD #: 15 = Saline and/or alkali soil (special management need?)

Code	Label
1	yes
2	no

FIELD #: 16 = Type of irrigation

Application of water to soils to assist in production of crops (fo 1982, or for at least 2 of the last 4 years)

Code	Label
0	Not irrigated
1	Gravity irrigation
2	Pressure irrigation
3	Gravity and pressure irrigation

FIELD #: 17 = Water source for irrigation

Code	Label
0	Not irrigated
1	Well
2	Pond, lake, or reservoir
3	Perennial stream
4	Lagoon or other waste water
5	Combination

FIELD #: 18 = Water provision; irrigation at least 1/2 of water requirements?

Irrigation provides at least 1/2 of the water requirements

Code	Label
1	yes
2	no

FIELD #: 19 = Flood prone area?

Code	Label
1	yes
2	no

FIELD #: 20 = Cover/use, general

General land cover/use based upon specific land cover/use and cropping history

Code	Label
A	Cropland
B	Pastureland and native pasture
C	Rangeland
D	Forest land
E	Other lands in farms
F	Barren lands
G	Other lands
H	Urban and built-up land
I	Rural transportations
J	Water (census)
K	Small water (non-census)

FIELD #: 21 = Cover/use, major

Code		Label
A		Cropland
	A	Cropland, horticulture
	B	Cropland, row crops
	C	Cropland, close grown crops
	E	Cropland, other
	F	Cropland, hayland
B		Pastureland and native pasture
	A	Pastureland and native pasture
C		Rangeland
D		Forest land
	A	Forest land (land stocked by forest trees, or bearing evidense of such tree cover and not currently developed for nonforest use)
E		Other lands in farms
	A	Other land in farms
F		Barren lands
	A	Barren land
G		Other lands
	A	Other lands
H		Urban and built-up land
I		Rural transportations

J	Water (census)
K	Small water (non-census)
A	Water body less than 40 acres
B	Small perennial stream

FIELD #: 22 = Cover/use, specific

Code	Label
A	Cropland
A	Cropland, horticulture
001	Fruit
002	Nut
003	Vineyard
004	Bush Fruit
005	Berries
006	Other horticulture
B	Cropland, row crops
011	Corn
012	Sorghum
013	Soybeans
014	Cotton
015	Peanuts
016	Tobacco
017	Sugarbeets
018	Potatoes
019	Other vegetables
020	All other row crops
021	Sunflowers
C	Cropland, close grown crops
111	Wheat
112	Oats
113	Rice
114	Barley
115	Flax
116	All other close grown crops
E	Cropland, other
120	Summer fallow
139	Aquaculture
140	Other cropland not planted
151	Cool season grass/hay (Grassland in rotation)
152	Warm season grass/hay (Grassland in rotation)
153	Legume/hay (Grassland in rotation)
154	Legume-grass/hay (Grassland in rotation)
221	Cool season grass (Grassland in rotation)
222	Warm season grass (Grassland in rotation)
223	Legume (Grassland in rotation)

224	Legume-grass mixed (Grassland in rotation)
225	Grass-forbs mixed (Grassland in rotation)
226	Grass-forbs-legume mixed (Grassland in rotation)
F	Cropland, hayland
151	Cool season grass/hay
152	Warm season grass/hay
153	Legume/hay
154	Legume-grass/hay
B	Pastureland and native pasture
A	Pastureland and native pasture
221	Cool season grass
222	Warm season grass
223	Legume
224	Legume-grass mixed
225	Grass-forbs mixed
226	Grass-forbs-legume mixed
C	Rangeland
250	Rangeland and tundra (native grasses, forbs, and shrubs valuable for forage)
D	Forest land
A	Forest land (land stocked by forest trees, or bearing evidense of such tree cover and not currently developed for nonforest use)
341	Forest land, grazed
342	Forest land, not grazed
E	Other lands in farms
A	Other land in farms
400	Farmsteads and ranch headquarters
401	Other land in farms
F	Barren lands
A	Barren land
611	Dry salt flats
612	Bare exposed rock
613	Strip mines, quarries, gravel, and borrow pits
614	Beaches
615	Sand dumes
616	Mixed barren lands
617	Mud flats
618	River wash
619	Oil wasteland
620	Other barren land
G	Other lands
A	Other lands
630	Permanent snow and ice fields
650	All other land
H	Urban and built-up land

- 710 Urban and built-up land, in units greater than 10 acres
- 730 Small built-up area (0.25-10 acres)
- I Rural transportations
  - 800 Rural transportation
- J Water (census)
  - 920 Water, Census (water bodies greater than 40 acres and perennial streams wider than 1/8 mile)
- K Small water (non-census)
  - A Water body less than 40 acres
    - 931 Water body 2-40 acres
    - 932 Water body less than 2 acres
  - B Small perennial stream
    - 941 Perennial stream less than 66 feet wide
    - 942 Perennial stream 66-660 feet wide

FIELD #: 23 = Use of the Land or Water

Land Use

Code	Label
10	Irrigation
20	Livestock
30	Water supply (municipal, industrial, household, etc.)
40	Aquaculture
50	Recreation
60	Fish and Wildlife
70	Erosion and sediment control
80	Other (flood prevention, water quality control, power, etc.)

Water Use

Code	Label
00	Not classified (urban and built-up areas, Census water, rural transportation
11	Crop production (harvested food, feed, forage, oil, horticulture, and fibre crops other than wood)
12	Livestock grazing
13	Wood production (refers to wood growth--not limited to wood that is harvested)
14	Idle
21	Residential
22	Commercial
23	Industrial
24	Institutional
25	Transmission (power, microwave, pipeline, etc.)
31	Waste disposal
41	Wilderness (designated)
42	Wildlife (designated)

43	Recreation (designated)
44	Nature study (designated)
51	Research and experimentation
61	Military
87	Other roads
99	None of the above

FIELD #: 24 = Cropping history 1981
Specific land/cover use for 1981

Code	Label
	Cropland, horticulture
001	Fruit
002	Nut
003	Vineyard
004	Bush Fruit
005	Berries
006	Other horticulture
	Cropland, row crops
011	Corn
012	Sorghum
013	Soybeans
014	Cotton
015	Peanuts
016	Tobacco
017	Sugarbeets
018	Potatoes
019	Other vegetables
020	All other row crops
021	Sunflowers
	Cropland, close grown crops
111	Wheat
112	Oats
113	Rice
114	Barley
115	Flax
116	All other close grown crops
	Cropland, other
120	Summer fallow
139	Aquaculture
140	Other cropland not planted
	Cropland, hayland
151	Cool season grass/hay
152	Warm season grass/hay
153	Legume/hay
154	Legume-grass/hay
	Pastureland and native pasture

221	Cool season grass
222	Warm season grass
223	Legume
224	Legume-grass mixed
225	Grass-forbs mixed
226	Grass-forbs-legume mixed
250	Rangeland and tundra (land on which the natural potentia plant coder is composed principally of native grasses, forbs, and shrubs valuable for forage)
	Forest land (land stocked by forest trees, or bearing evidense of such tree cover and not currently developed for nonforest use)
341	Forest land, grazed
342	Forest land, not grazed
	Other land in farms
400	Farmsteads and ranch headquarters
401	Other land in farms
	Barren land
611	Dry salt flats
612	Bare exposed rock
613	Strip mines, quarries, gravel, and borrow pits
614	Beaches
615	Sand dumes
616	Mixed barren lands
617	Mud flats
618	River wash
619	Oil wasteland
620	Other barren land
	Other lands
630	Permanent snow and ice fields
650	All other land
710	Urban and built-up land, in units greater than 10 acres
730	Small built-up area (0.25-10 acres)
800	Rural transportation
920	Water, Census (water bodies greater than 40 acres and perennial streams wider than 1/8 mile)
	Water body less than 40 acres
931	Water body 2-40 acres
932	Water body less than 2 acres
	Small perennial stream
941	Perennial stream less than 66 feet wide
942	Perennial stream 66-660 feet wide

FIELD #: 25 = Cropping history 1980
Specific land/cover use for 1980

Code	Label
	Cropland, horticulture
001	Fruit
002	Nut
003	Vineyard
004	Bush Fruit
005	Berries
006	Other horticulture
	Cropland, row crops
011	Corn
012	Sorghum
013	Soybeans
014	Cotton
015	Peanuts
016	Tobacco
017	Sugarbeets
018	Potatoes
019	Other vegetables
020	All other row crops
021	Sunflowers
	Cropland, close grown crops
111	Wheat
112	Oats
113	Rice
114	Barley
115	Flax
116	All other close grown crops
	Cropland, other
120	Summer fallow
139	Aquaculture
140	Other cropland not planted
	Cropland, hayland
151	Cool season grass/hay
152	Warm season grass/hay
153	Legume/hay
154	Legume-grass/hay
	Pastureland and native pasture
221	Cool season grass
222	Warm season grass
223	Legume
224	Legume-grass mixed
225	Grass-forbs mixed
226	Grass-forbs-legume mixed

250	Rangeland and tundra (land on which the natural potentia plant coder is composed principally of native grasses, forbs, and shrubs valuable for forage)
	Forest land (land stocked by forest trees, or bearing evidense of such tree cover and not currently developed for nonforest use)
341	Forest land, grazed
342	Forest land, not grazed
	Other land in farms
400	Farmsteads and ranch headquarters
401	Other land in farms
	Barren land
611	Dry salt flats
612	Bare exposed rock
613	Strip mines, quarries, gravel, and borrow pits
614	Beaches
615	Sand dumes
616	Mixed barren lands
617	Mud flats
618	River wash
619	Oil wasteland
620	Other barren land
	Other lands
630	Permanent snow and ice fields
650	All other land
710	Urban and built-up land, in units greater than 10 acres
730	Small built-up area (0.25-10 acres)
800	Rural transportation
920	Water, Census (water bodies greater than 40 acres and perennial streams wider than 1/8 mile)
	Water body less than 40 acres
931	Water body 2-40 acres
932	Water body less than 2 acres
	Small perennial stream
941	Perennial stream less than 66 feet wide
942	Perennial stream 66-660 feet wide

FIELD #: 26 = Cropping history 1979

Specific land/cover use for 1979

Code	Label
	Cropland, horticulture
001	Fruit
002	Nut
003	Vineyard

004	Bush Fruit
005	Berries
006	Other horticulture
	Cropland, row crops
011	Corn
012	Sorghum
013	Soybeans
014	Cotton
015	Peanuts
016	Tobacco
017	Sugarbeets
018	Potatoes
019	Other vegetables
020	All other row crops
021	Sunflowers
	Cropland, close grown crops
111	Wheat
112	Oats
113	Rice
114	Barley
115	Flax
116	All other close grown crops
	Cropland, other
120	Summer fallow
139	Aquaculture
140	Other cropland not planted
	Cropland, hayland
151	Cool season grass/hay
152	Warm season grass/hay
153	Legume/hay
154	Legume-grass/hay
	Pastureland and native pasture
221	Cool season grass
222	Warm season grass
223	Legume
224	Legume-grass mixed
225	Grass-forbs mixed
226	Grass-forbs-legume mixed
250	Rangeland and tundra (land on which the natural potentia plant coder is composed principally of native grasses, forbs, and shrubs valuable for forage)
	Forest land (land stocked by forest trees, or bearing evidense of such tree cover and not currently developed for nonforest use)
341	Forest land, grazed
342	Forest land, not grazed

	Other land in farms
400	Farmsteads and ranch headquarters
401	Other land in farms
	Barren land
611	Dry salt flats
612	Bare exposed rock
613	Strip mines, quarries, gravel, and borrow pits
614	Beaches
615	Sand dunes
616	Mixed barren lands
617	Mud flats
618	River wash
619	Oil wasteland
620	Other barren land
	Other lands
630	Permanent snow and ice fields
650	All other land
710	Urban and built-up land, in units greater than 10 acres
730	Small built-up area (0.25-10 acres)
800	Rural transportation
920	Water, Census (water bodies greater than 40 acres and perennial streams wider than 1/8 mile)
	Water body less than 40 acres
931	Water body 2-40 acres
932	Water body less than 2 acres
	Small perennial stream
941	Perennial stream less than 66 feet wide
942	Perennial stream 66-660 feet wide

FIELD #: 27 = Double-cropping used?

Code	Label
1	yes
2	no

FIELD #: 28 = Conservation practice 1

Conservation practices that are currently in use on the land -- up to three practices can be identified.

Code	Label
000	None of the above, or not applicable
312	Waste management system
314	Brush management
329	Conservation tillage system
330	Contour farming
362	Diversion

380	Farmstead and feedlot windbreak
392	Field windbreak
410	Grade stabilization structures
412	Grassed waterway or outlet
428	Irrigation water conveyance, ditch and canal lining
430	Irrigation water conveyance, pipeline
449	Irrigation water management
510	Pasture and hayland management
528	Proper grazing use
543	Land reconstruction, abandoned mined land
544	Land reconstruction, currently mined land
550	Range seeding
585	Stripcropping, contour
589	Stripcropping, wind
600	Terrace
606	Subsurface drain
607	Surface drainage, field ditch
608	Surface drainage, main or lateral
612	Tree planting
666	Woodland improvement

FIELD #: 29 = Conservation practice 2

Conservation practices that are currently in use on the land -- up to three practices can be identified.

Code	Label
000	None of the above, or not applicable
312	Waste management system
314	Brush management
329	Conservation tillage system
330	Contour farming
362	Diversion
380	Farmstead and feedlot windbreak
392	Field windbreak
410	Grade stabilization structures
412	Grassed waterway or outlet
428	Irrigation water conveyance, ditch and canal lining
430	Irrigation water conveyance, pipeline
449	Irrigation water management
510	Pasture and hayland management
528	Proper grazing use
543	Land reconstruction, abandoned mined land
544	Land reconstruction, currently mined land
550	Range seeding
585	Stripcropping, contour
589	Stripcropping, wind

600	Terrace
606	Subsurface drain
607	Surface drainage, field ditch
608	Surface drainage, main or lateral
612	Tree planting
666	Woodland improvement

FIELD #: 30 = Conservation practice 3

Conservation practices that are currently in use on the land -- up to three practices can be identified.

Code	Label
000	None of the above, or not applicable
312	Waste management system
314	Brush management
329	Conservation tillage system
330	Contour farming
362	Diversion
380	Farmstead and feedlot windbreak
392	Field windbreak
410	Grade stabilization structures
412	Grassed waterway or outlet
428	Irrigation water conveyance, ditch and canal lining
430	Irrigation water conveyance, pipeline
449	Irrigation water management
510	Pasture and hayland management
528	Proper grazing use
543	Land reconstruction, abandoned mined land
544	Land reconstruction, currently mined land
550	Range seeding
585	Stripcropping, contour
589	Stripcropping, wind
600	Terrace
606	Subsurface drain
607	Surface drainage, field ditch
608	Surface drainage, main or lateral
612	Tree planting
666	Woodland improvement

FIELD #: 31 = Conservation treatment needed

Code	Label
00	Not applicable
01	Adequately protected
02	Erosion control needed
03	Drainage needed

04	Irrigation management needed
05	Forage needs protection only
06	Forage needs improvement with no brush management
07	Forage needs improvement with brush management
08	Forage reestablishment with no brush management
09	Forage reestablishment with brush management
10	Establishment and reinforcement of timber
11	Timber stand improvement
12	Conservation treatment needed to improve timber crops
13	Treatment not feasible

FIELD #: 32 = K-Factor
Soil erodibility factor for Universal Soil Loss Equation (USLE)
UNITS:
MISC: decimal suppressed

FIELD #: 33 = R-Factor
Rainfall factor for Universal Soil Less Equation (USLE)
UNITS:

FIELD #: 34 = C-Factor
Cropping-management factor for Universal Soil Loss Equation (USLE)
UNITS:
MISC: decimal suppressed

FIELD #: 35 = P-Factor
Erosion control practice factor for Universal Soil Loss Equation (USLE)
UNITS:
MISC: decimal suppressed

FIELD #: 36 = Slope length for Universal Soil Loss Equation (USLE)
UNITS: feet
MISC:

FIELD #: 37 = Slope percent for Universal Soil Loss Equation (USLE)
UNITS: %
MISC: decimal suppressed

FIELD #: 38 = Universal Soil Loss Equation (USLE)
USLE calculation of estimated average soil movement due to sheet a

Code	Label
0	None, or not applicable
1	Modified LS, for frozen soil
2	Version for R-1 and R-2 areas

FIELD #: 39 = Universal Soil Loss Equation (USLE) flag
Modified version of USLE applied
UNITS:
MISC:

FIELD #: 40 = Universal Soil Loss Equation (USLE) tons
Estimated average annual tons of soil movement due to sheet and rill erosion
UNITS: 100s of tons per yea
MISC:

FIELD #: 41 = Wind erosion
Estimated average annual soil movement due to wind erosion
UNITS: 100s of tons per yea
MISC: decimal suppressed

FIELD #: 42 = Wind erosion tons
Estimated average annual soil movement due to wind erosion
UNITS: 100s of tons per yea
MISC:

FIELD #: 43 = Dominant soil and water problem
Dominant soil and water related reason inhibiting or preventing conversion of land to cropland

Code	Label
00	None
01	Common Flooding
02	Very low fertility
03	Very stony or rocky surface
04	Very high erosion potential
05	Lack of dependable irrigation water supply
06	No water available for irrigation
07	Wetland types (1-20)
08	Saline and/or alkali
09	Very Droughty (very low available water)
10	Restrictive root zone
11	Wetness problem (South Carolina only)
21	Short growing season
99	Currently cropland, built-up, transportation, or water; or is 7E, 7W, 7S, or class 8 soil

FIELD #: 44 = Secondary soil and water problem
Secondary soil and water related reason inhibiting or preventing conversion of land to cropland

Code	Label
00	None
01	Common Flooding
02	Very low fertility
03	Very stony or rocky surface
04	Very high erosion potential
05	Lack of dependable irrigation water supply
06	No water available for irrigation
07	Wetland types (1-20)
08	Saline and/or alkali
09	Very Droughty (very low available water)
10	Restrictive root zone
11	Wetness problem (South Carolina only)
21	Short growing season
99	Currently cropland, built-up, transportation, or water; or is 7E, 7W, 7S, or class 8 soil

FIELD #: 45 = Dominant non-soil or water problem

Dominant other (than soil or water) reason that would inhibit or prevent conversion to cropland

Code	Label
00	None--no reason why land cannot be converted to cropland
11	Land being held for urban or related development
12	Small tract--too small for efficient agricultural operation as a farm
13	Isolated tract--of sufficient size for efficient use of modern machinery but too distant from other farmland or too inaccessible for incorporation into efficient farm unit
14	Landowner committed to noncropland uses
99	Currently cropland, built-up, transportation, or water; or is 7E, 7W, 7S, or class 8 soil

FIELD #: 46 = Type of effort necessary for conversion to cropland

Code	Label
01	None-can convert by beginning tillage
02	On-farm - can convert through actions by individual farmers
03	Multi-farm - informal or formal cooperation between neighbors to install systems
04	Project action required
09	Not applicable, e.g., zero potential
99	Currently cropland, built-up, transportation, or water; or 7E, 7W, 7S or class 8 land

FIELD #: 47 = Conversion potential

Potential for conversion to cropland within the foreseeable future

Code	Label
00	Zero potential
01	Conversion unlikely in forseeable future
02	Medium potential
03	High potential
99	Currently cropland, built-up, transportation, or water; or is 7E, 7W, 7S, or class 8 soil

FIELD #: 48 = Type of wetland

Code	Label
00	None
01	Seasonally flooded basins or flats Few inches in upland; few feet along rivers
02	Inland fresh meadows Few inches after heavy rains
03	Inland shallow fresh marshes Up to 6 inches
04	Inland deep fresh marshes Up to 3 feet
05	Inland open fresh water Up to 10 feet; marshy border may be present
06	Shrub swamps Up to 6 inches
07	Wooded swamps Up to 1 foot
08	Bogs Shallow ponds may be present
09	Inland saline flats Few inches after heavy rain
10	Inland saline marshes Up to 3 feet
11	Inland open saline water Up to 10 feet; marshy border
12	Coastal shallow fresh marshes Up to 6 inches at high tide
13	Coastal deep fresh marshes Up to 3 feet at high tide
14	Coastal open fresh water Up to 10 feet; marshy border often present
15	Coastal salt flats May have few inches at high tide
16	Coastal salt meadows

	May have few inches at high tide
17	Irregularly flooded salt marshes Few inches at high tide
18	Regularly flooded salt marshes Up to 1 foot at high tide
19	Sounds and bays Up to 10 feet at high tide
20	Mangrove swamps Up to 3 feet

FIELD #: 49 = Kind of wetland vegetation

Code	Label
0	None
1	Emergent
2	Scrub/shrub
3	Forested

FIELD #: 50 = Kind of wetland system

A complex of wetland and deep water habitats influenced by hydrologic, geomorphological, chemical, and/or biological factors

Code	Label
0	No kind
1	Marine
2	Estuarine
3	Riverine
4	Lacustrine
5	Palustrine

FIELD #: 51 = Riparian area kind

Code	Label
0	None
1	Natural Streambank
2	Manmade canal or ditch bank
3	Natural pond or lak shoreline
4	Manmade pond or reservoir shoreline
5	Tidal area shoreline

FIELD #: 52 = Riparian vegetation kind

Code	Label
0	None
1	Trees
2	Shrubs

3	Forbs
4	Grass and grasslike plants
5	Mixed
6	Other

FIELD #: 53 = Riparian vegetation; width of strip

Code	Label
0	None
1	Less than 100 feet
2	100-500 feet
3	Greater than 500 feet

FIELD #: 54 = Distance to cropland
Distance from point to nearest occurrence of cropland (in feet); i
'9999' is nearest occurence is further than 5,280 feet
UNITS: feet
MISC:

FIELD #: 55 = Distance to forest land
Distance from point to nearest occurrence of forest land (in feet)
is '9999' is nearest occurence is further than 5,280 feet
UNITS: feet
MISC:

FIELD #: 56 = Distance to grassland
Distance from point to nearest occurrence of pastureland or
rangeland (in feet); is '9999' is nearest occurence is further tha
5,280 feet
UNITS: feet
MISC:

FIELD #: 57 = Distance to water
Distance from point to nearest occurrence of water (in feet); is
'9999' is nearest occurence is further than 5,280 feet
UNITS: feet
MISC:

FIELD #: 58 = Distance to wetlands
Distance from point to nearest occurrence of wetlands (in feet); i
'9999' is nearest occurence is further than 5,280 feet
UNITS: feet
MISC:

FIELD #: 59 = Distance to built-up land
Distance from point to nearest occurrence of farmsteads, urban and

built-up, roads, etc. (in feet); is '9999' is nearest occurence is further than 5,280 feet
UNITS: feet
MISC:

FIELD #: 60 = Winter cover
Winter ground cover of the last harvested crop

Code	Label
0	None, or not cropland
1	Live crop
2	Cropland residue

FIELD #: 61 = Winter cover height

Height of crop or residue remaining over winter
UNITS: inches
MISC:

FIELD #: 62 = Residue remains upright over winter?

Code	Label
0	Not cropland
1	Yes
2	No

FIELD #: 63 = Pastureland condition rating

Code	Label
0	Not pastureland
2	Good
3	Fair
4	Poor
9	Not applicable

FIELD #: 64 = Woody canopy cover, for pastureland
Canopy cover of woody plants, if pastureland or native pasture

Code	Label
0	Not pastureland
1	0-9%
2	10-25%
3	26-55%
4	56-100%

FIELD #: 65 = Rangeland condition rating (percent climax vegetation)

Code	Label
0	Not rangeland
1	Excellent (76-100%)
2	Good (51-75%)
3	Fair (26-50%)
4	Poor (0-25%)
8	Annual range
9	Not applicable

FIELD #: 66 = Woody canopy cover, for rangeland

Code	Label
0	Not rangeland
1	0-9%
2	10-25%
3	26-55%
4	56-100%

FIELD #: 67 = Rangeland condition trend

Apparent trend in condition on the soil and/or vegetation resource for rangeland

Code	Label
0	Not rangeland
1	Up (soil and/or vegetation resources are improving)
2	Even (not readily apparent)
3	Down (resources are deteriorating)

FIELD #: 68 = Grazing level, for rangeland

Code	Label
0	Not rangeland
1	Not routinely grazed
2	Routinely grazed but presently deferred
3	Currently grazed, lightly to properly used
4	Currently grazed, excessively used

FIELD #: 69 = Forest type, general category

Code	Label
A	White-red-jack pine
B	Spruce-fir
C	Longleaf-splash pine
D	Loblolly-shortleaf pine
E	Oak-pine
F	Oak-hickory

G	Oak-gum-cypress
H	Elm-ash-cottonwood
I	Maple-beech-birch
J	Aspen-birch
K	Low productivity type
L	Tropical forest
M	Nonstocked
N	Douglas-fir
O	Ponderosa pine
P	Western white pine
Q	Fir-spruce
R	Hemlock-Sitka spruce
S	Larch
T	Lodgepole pine
U	Redwood
V	Hardwoods
W	Other conifers
X	Savanna
Y	Alaskan species

FIELD #: 70 = Forest type, specific category

Code	Label
A	White-red-jack pine
001	Jack pine
002	Red pine
003	White pine
004	White pine-hemlock
005	Hemlock
B	Spruce-fir
011	Balsam fir
012	Black spruce
013	Red spruce-Balsam fir
014	Northern white-cedar
015	Tammarack
016	White spruce
C	Longleaf-splash pine
021	Longleaf pine
022	Slash pine
D	Loblolly-shortleaf pine
031	Loblolly pine
032	Shortleaf pine
033	Virginia pine
034	Sand pine
035	Eastern redcedar
036	Pond pine

037	Spruce pine
038	Pitch pine
039	Table-mountain pine
E	Oak-pine
041	White pine-northern red oak-white ash
042	Eastern redcedar-hardwood
043	Longleaf pine-scrub oak
044	Shortleaf pine-oak
045	Virginia pine-southern red oak
046	Loblolly pine-hardwood
047	Slash pine-hardwood
049	Other oak-pine
F	Oak-hickory
051	Post oak, black oak or bear oak
052	Chestnut oak
053	White oak-red oak-hickory
054	White oak
055	Northern red oak
056	Yellow poplar-white oak-northern red oak
057	Southern scrub oak
058	Sweetgum-yellow poplar
059	Mixed hardwoods
G	Oak-gum-cypress
061	Swamp chestnut oak-cherry bark oak
062	Sweetgum-Nuttall oak-willow oak
063	Sugarberry-American elm-green ash
065	Overcup oak-water hickory
066	Atlantic white cedar
067	Baldcypress-water tupelo
068	Sweetbay-swamp tupelo-red maple
H	Elm-ash-cottonwood
071	Black ash-American elm-red maple
072	River birch-sycamore
073	Cottonwood
074	Willow
075	Sycamore-pecan-American elm
076	Silver maple-American elm
I	Maple-beech-birch
081	Sugar maple-beech-yellow birch
J	Aspen-birch
091	Aspen
092	Paper birch
K	Low productivity type
095	Unproductive forest
L	Tropical forest
097	Tropical forest

M	Nonstocked
099	Nonstocked forest land
N	Douglas-fir
101	Douglas-fir
102	Douglas-fir-Western hemlock
103	Port orford-cedar-Douglas-fir
O	Ponderosa pine
111	Ponderosa pine
112	Jeffrey pine
113	Ponderosa pine-sugar pine-fir
114	Bishop pine-Monterey pine
P	Western white pine
121	Western white pine
Q	Fir-spruce
131	White fir
132	Red fir
134	Pacific silver fire-hemlock
135	Engelmann spruce
136	Engelmann spruce-subalpine fir
R	Hemlock-Sitka spruce
141	Western red cedar
142	Sitka spruce
147	Mountain hemlock-subalpine fir
148	Western Hemlock
S	Larch
155	Larch-Douglas-fir
156	Grand fir-larch-Douglas-fir
157	Ponderosa pine-larch-Douglas fir
T	Lodgepole pine
161	Lodgepole pine
162	Shore pine
U	Redwood
171	Redwood
V	Hardwoods
175	Red alder-big leaf maple
176	Poplar-birch
177	Aspen
178	California balck oak
179	Cottonwood-willow, tamarish (California only)
180	Canyon live oak
181	Oak-madrone (Tanoak-California live oak-Pacific madrone)
182	Chaparral (Quercus-silktasole-ceanothus-manzanita-chamaise mountain mahogony-sumac-mesquite) Kiobe (Hawaii only)
183	Ohia
184	Oregon white oak
185	Interior live oak

186	Eucalyptus
W	Other conifers
191	Coulter pine
193	Pinyon-juniper
194	Knobcone pine
195	Bristlecone pine
196	White pine
197	Limber pine
X	Savanna
205	Digger pine-oak-blue oak
Y	Alaskan species
211	White spruce
212	White spruce-birch
213	Black spruce

FIELD #: 71 = Canopy cover of the trees, for forest land

Code	Label
0	Not forest land
1	0-9%
2	10-25%
3	26-55%
4	56-100%

FIELD #: 72 = Basal area/stem count

Basal area of the stand (in square feet per acre), if average DBH is at least 5 inches; if DBH is less than 5 inches, then stocking based on stem count use is given.

Code	Label
000	Nonstocked, or not forest land
900	Poorly stocked
901	Moderatley stocked
902	Fully stocked

FIELD #: 73 = Diameter at Breast Height (DBH)

Average Diameter at breast height (DBH), for forest land inches

UNITS:

MISC:

FIELD #: 74 = Forest understory composition

Primary plant group for forest land understory composition

Code	Label
0	None

1	Woody
2	Forbs
3	Grass and grass-like plants

FIELD #: 75 = Understory forage value

Forage value rating of understory for forest land, based upon percent of understory production by preferred species

Code	Label
0	Not forest land
1	Very high (51-100%)
2	High (31-50%)
3	Moderate (11-30%)
4	Low (0-10%)
9	Not applicable, i.e., not suitable for grazing

FIELD #: 76 = Soils-5

Soils-5 identification block

Field Column(s)	Description of Field
1-6	SCS-Soils-5 Interpretations Record Number (alpha-numeric)
7-9	Surface layer texture modifier, as per Table 603-45 on page 603-198 of the National Soils Handbook (July 1983) (alpha)
10-13	Surface layer texture or term used in-lieu of texture, as per Table 603-45 of the National Soils Handbook (July 1983) (alpha)
14-15	Slope class range, lower limit (numeric)
16-17	Slope class range, upper limit (numeric)
18	Flooding class: N = none R = rare O = occasional F = frequent
19-20	Capability and predicted yields for crops and pasture-- class determining phase number, indicating which portion (or line) of this section should be used (numeric: 00-12)
21-22	Woodland suitability--phase number indicating which portion (line) should be used (numeric: 00-15)
23	Windbreaks, species and heights--phase number indicating which portion should be used (numeric: 0-2)
24	Wildlife habitat suitability--phase number (numeric: 0-6)
25	Rangeland potential native plant community--phase number (numeric: 0-5)
26	Forest understory vegetation, potential native plant community--phase number (numeric: 0-5)
27	Septic tank absorption field--phase number for use interpretations (numeric: 0-5)

28 Sewage lagoons--phase number for use interpretations (numeric: 0-5)
29 Sanitary landfill (trench)--phase number of use interpretations (numeric: 0-5)
30 Sanitary landfill (area)--phase number for use interpretations (numeric: 0-5)
31 Daily cover for landfill--phase number for use interpretations (numeric: 0-5)
32 Shallow excavations--phase number for use interpretations (numeric: 0-5)
33 Dwellings without basements--phase number for use interpretations (numeric: 0-5)
34 Dwellings with basements--phase number for use interpretations (numeric: 0-5)
35 Small commercial buildings--phase number for use interpretations (numeric: 0-5)
36 Local streets and roads--phase number for use interpretations (numeric: 0-5)
37 Lawns, landscaping, and golf fairways--phase number for use interpretations (numeric: 0-5)
38 Roadfill--phase number for use interpretations (numeric: 0-5)
39 Sand--phase number for use interpretations (numeric: 0-5)
40 Gravel--phase number for use interpretations (numeric: 0-5)
41 Topsoil--phase number for use interpretations (numeric: 0-5)
42 Pond reservoir area--phase number for use interpretations (numeric: 0-5)
43 Embankments, dikes, and levees--phase number for use interpretations (numeric: 0-5)
44 Excavated ponds that are aquifer fed--phase number for use interpretations (numeric: 0-5)
45 Drainage--phase number for use interpretations (numeric: 0-5)
46 Irrigation--phase number for use interpretations (numeric: 0-5)
47 Terraces and diversions--phase number for use interpretations (numeric: 0-5)
48 Grassed waterways--phase number for use interpretations (numeric: 0-5)
49 Camp areas--phase number for use interpretations (numeric: 0-5)
50 Picnic areas--phase number for use interpretations (numeric: 0-5)
51 Playgrounds--phase number for use interpretations (numeric: 0-5)
52 Paths and trails--phase number for use interpretations (numeric: 0-5)

```
+-----------------+
|       NRI       |
| "Dummy" Records |
+-----------------+
```

Provided together with the sample observations you have extracted from the 1982 NRI database are two types of "dummy" records. These "dummy" records are an essential component of the NRI database; they are needed to account for land areas not physically sampled by the NRI. The two categories of "dummy" records are:

(1) Dummy records for urban and built-up areas, roads, and bodies of water. This category includes "dummy" observations for any land area associated with a land cover/use code (Field 22) greater than 700. A single "dummy" observation exists for each MLRA within a county area for each type of applicable land cover/use. These observations have a blank in the PSU

number field (Field 2) rather than a number; many other fields are also blank or "zeroed out." The expansion factor (Field 7) for these observations represents the total acreage under the particular land cover/use within the specified "MLRA within a county" area. Of the 841,860 records in the 1982 NRI database, 42,521 are "dummy" records of this category.

(2) Non-sampled farmstead areas. Farmstead acres, associated with a land cover/use code (Field 22) of 400, are handled differently from the first category; there exist both "dummy" observations and specific sample point observations containing '400' in Field 22. Dummy farmstead points have Fields 2 and 3 "zeroed out" but have useable codes in all other appropriate fields. There are 2,264 of these records in the complete 1982 NRI database.

NOTE: Some Land Cover/Use codes will not retrieve any sample point records, they will only retrieve non-sample point data.

Land Cover/Use specific code	Type of data retrieved
100-300, 401, 600	sample point, only
400	sample point and non-sample point
700-942	non-sample point, only

```
+----------------< Output Totals >-----------------+
  Total # Sample point records extracted = 298
  Total # Farmstead records extracted = 8
  Total # County records extracted = 327
+--------------------------------------------------+
```

Appendix E: National Level Presentations of INFeRS

ASIS — Katherine Chiang	May 1989
CODATA — Katherine Chiang	July 1990
ALA/STS Section — Marijo Wilson	June 1990
Computers in Libraries — Leslie McLane	September 1990
USAIN — Marijo Wilson	November 1990

Appendix F: Handouts from presentations of INFeRS

Originally the system was called Aqueduct, all these handouts refer to INFeRS by that name.

From Mann Library

AGRICULTURAL DATA ONLINE

Use your own computer, dial into the Mann Library computer, select the data you need, run a search, and transfer the datafile to your computer.

It's easy to use.

AVAILABLE NOW

National Resources Inventory

1982 Inventory of non-Federal lands, 800,000 records describing land use, erosion, soil and water types.

COMING SOON

USDA Crop Reporting Board statistics

Production statistics for 67 crops, county level, 20 years.

Commodity trading data

Trading history for major commodities, in some cases as far back as 1968.

If you would like online access to these statistics please complete the attached sheet and return it to Computer Files, Mann Library. You will be sent instructions.

If you have questions please contact Katherine Chiang, Computer Files 255-2199, anmj@cornella.

To Mann Library Computer Files Service
Please send me instructions on accessing data online from the Mann Library computer.

Name ______________________________

Campus Address ______________________________
____

Telephone number ______________ CIT computer id/e-mail address ______________

Status (check one)
______ Lecturer
______ Research Associate
______ Extension Associate
______ Assistant Professor
______ Associate Professor
______ Professor
______ Administrator
______ Support Staff
______ Graduate Student MS or MA
______ Graduate Student MS/Ph.D.
______ Graduate Student Ph.D.
______ Undergraduate Student

Department ______________________________

Which computers do you use? (Check any that apply)

______ Microcomputer: IBM and IBM compatible
______ Microcomputer: Apple Macintosh
______ Microcomputer: Other (please specify) ______________________________
______ Minicomputer (shared systems) which serve a department or limited work-group
______ Mainframe or minicomputer available to large segments of the Cornell community
______ Other (please specify) ______________________________

For any of the personal computers at your home, Cornell office or laboratory equipped with a modem or other device for communicating with other computers, check any that apply:
______ Telephone modem
______ Sytek network connection
______ Connection, or gateway access, to the campus backbone (CITNet, Internet)

What communications software are you using?

Macintosh
______ MacTerminal
______ VersaTerm
______ Red Tyder
______ Comet (Cornell Telnet)
______ NCSA Telnet
______ Other (please specify)

PC
______ Kermit
______ PC-Talk
______ ProComm
______ Cornell Telnet
______ Other (please specify)

BETA TESTERS

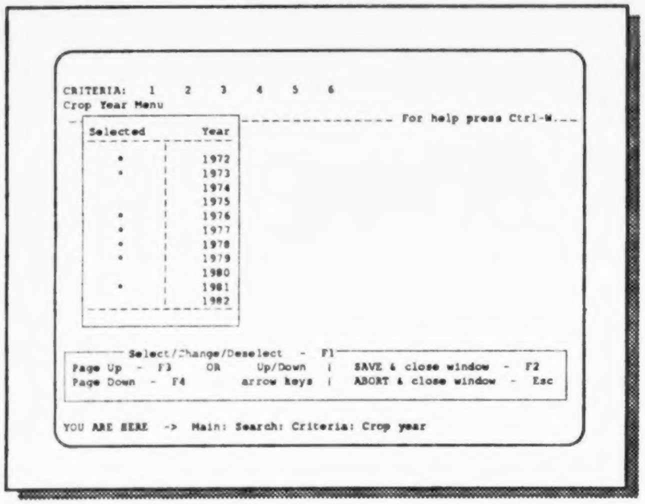
We are looking for a few faculty, graduate students, and undergraduates interested in helping us to test and critique the new service. We need approximately three hours of your time: a two hour session in the library, and after you have run some searches, an interview in your office or lab. You will learn how to retrieve data, while giving us valuable information we will use to improve the service.

To make it easy for you to use the service in your office after the test, we will come to your office, if necessary, and give you technical assistance in logging in from your computer.

We need your help. If you are interested in volunteering please check this box ☐

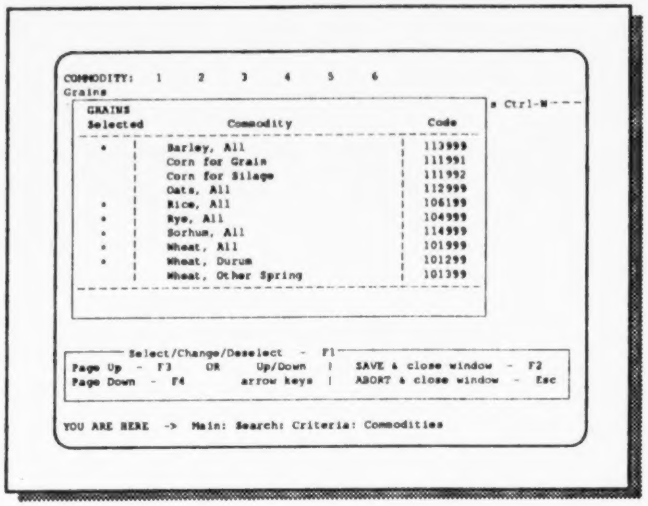
Please return this form to Katherine Chiang, Computer Files, Mann Library.

5



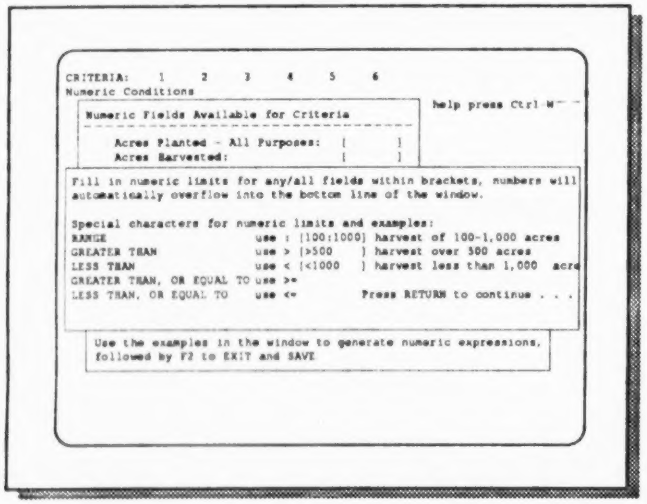
```
CRITERIA:   1    2    3    4    5    6
Crop Year Menu
 -  Selected       Year  --------------- For help press Ctrl-W.--
    ------------------
          •    |   1972
          •    |   1973
               |   1974
               |   1975
          •    |   1976
          •    |   1977
          •    |   1978
          •    |   1979
               |   1980
          •    |   1981
               |   1982
    ------------------

 ------ Select/Change/Deselect  -  F1 ---------------------------------
 Page Up   -  F3    OR    Up/Down    |   SAVE & close window  -  F2
 Page Down  -  F4        arrow keys  |   ABORT & close window  -  Esc

YOU ARE HERE  ->  Main: Search: Criteria: Crop year
```

6



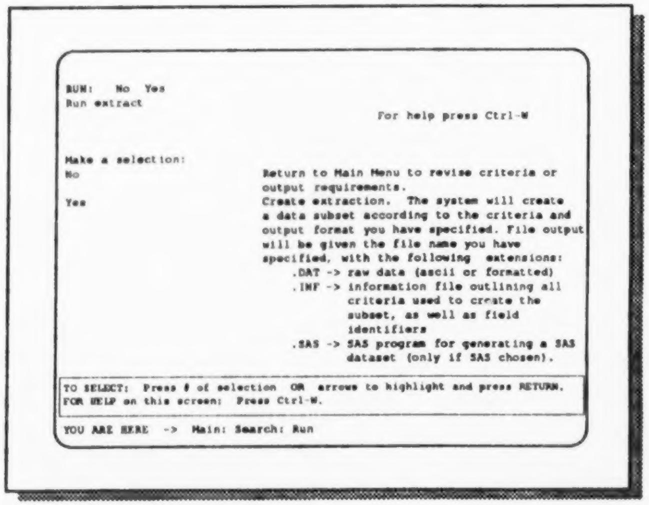
```
COMMODITY:  1    2    3    4    5    6
Grains
   GRAINS                                            s Ctrl-W---
   Selected          Commodity                 Code
  ----------------------------------------------------
      •    |     Barley, All               |  113999
           |     Corn for Grain            |  111991
           |     Corn for Silage           |  111992
           |     Oats, All                 |  112999
      •    |     Rice, All                 |  106199
      •    |     Rye, All                  |  104999
      •    |     Sorhum, All               |  114999
      •    |     Wheat, All                |  101999
      •    |     Wheat, Durum              |  101299
           |     Wheat, Other Spring       |  101399
  ----------------------------------------------------

 ------ Select/Change/Deselect  -  F1 ---------------------------------
 Page Up   -  F3    OR    Up/Down    |   SAVE & close window  -  F2
 Page Down  -  F4        arrow keys  |   ABORT & close window  -  Esc

YOU ARE HERE  ->  Main: Search: Criteria: Commodities
```

7



```
CRITERIA:   1    2    3    4    5    6
Numeric Conditions
  Numeric Fields Available for Criteria             help press Ctrl-W--
  ---------------------------------------
      Acres Planted - All Purposes:  [       ]
      Acres Harvested:               [       ]
Fill in numeric limits for any/all fields within brackets, numbers will
automatically overflow into the bottom line of the window.

Special characters for numeric limits and examples:
RANGE                       use : [100:1000] harvest of 100-1,000 acres
GREATER THAN                use > [>500    ] harvest over 500 acres
LESS THAN                   use < [<1000   ] harvest less than 1,000  acre
GREATER THAN, OR EQUAL TO use >=
LESS THAN, OR EQUAL TO      use <=        Press RETURN to continue . . .

   Use the examples in the window to generate numeric expressions,
   followed by F2 to EXIT and SAVE
```

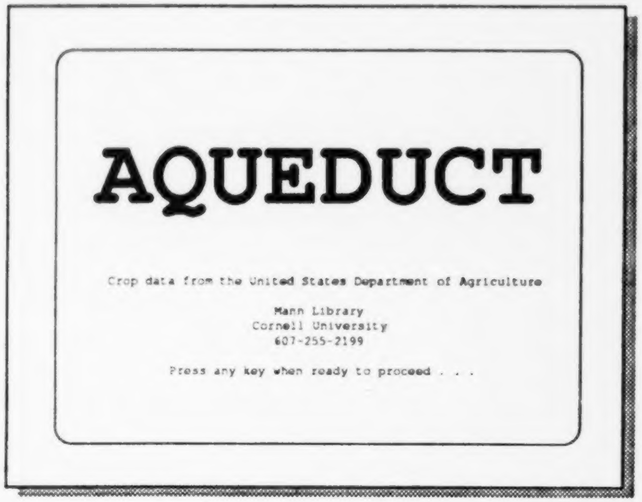
8



```
RUN:   No  Yes
Run extract
                                         For help press Ctrl-W


Make a selection:
No                           Return to Main Menu to revise criteria or
                             output requirements.
Yes                          Create extraction.  The system will create
                             a data subset according to the criteria and
                             output format you have specified. File output
                             will be given the file name you have
                             specified, with the following  extensions:
                                 .DAT -> raw data (ascii or formatted)
                                 .INF -> information file outlining all
                                         criteria used to create the
                                         subset, as well as field
                                         identifiers
                                 .SAS -> SAS program for generating a SAS
                                         dataset (only if SAS chosen).
TO SELECT:  Press # of selection  OR  arrows to highlight and press RETURN.
FOR HELP on this screen:  Press Ctrl-W.
YOU ARE HERE  ->  Main: Search: Run
```

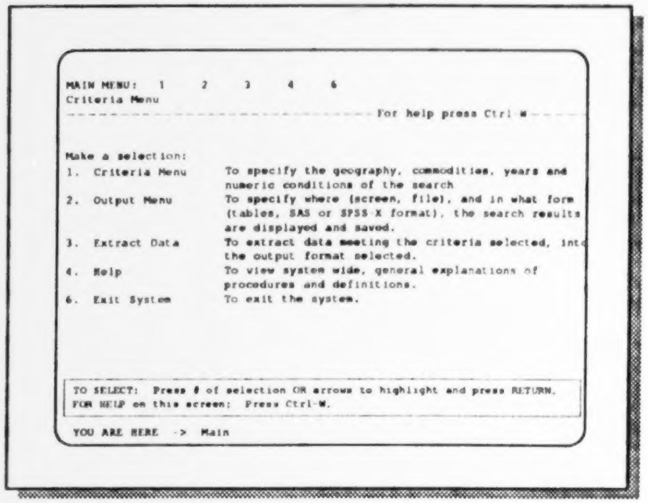
BEST COPY AVAILABLE

1



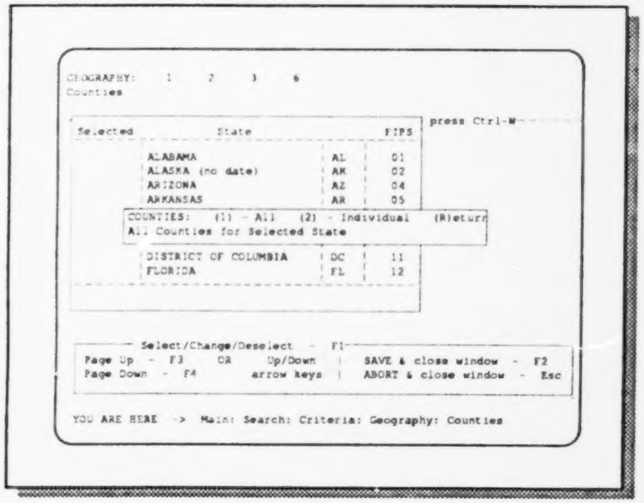
```
AQUEDUCT

Crop data from the United States Department of Agriculture

                      Mann Library
                   Cornell University
                      607-255-2199

            Press any key when ready to proceed . . .
```

2



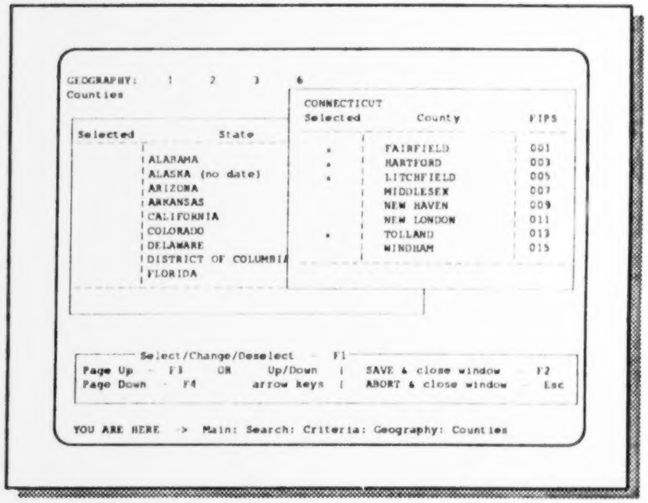
```
MAIN MENU:   1     2     3     4     6
Criteria Menu
---------------------------------------------For help press Ctrl-W-----

Make a selection:
1.  Criteria Menu      To specify the geography, commodities, years and
                       numeric conditions of the search
2.  Output Menu        To specify where (screen, file), and in what form
                       (tables, SAS or SPSS-X format), the search results
                       are displayed and saved.
3.  Extract Data       To extract data meeting the criteria selected, int
                       the output format selected.
4.  Help               To view system wide, general explanations of
                       procedures and definitions.
6.  Exit System        To exit the system.

 TO SELECT:  Press # of selection OR arrows to highlight and press RETURN.
 FOR HELP on this screen:  Press Ctrl-W.

 YOU ARE HERE  ->  Main
```

3



```
GEOGRAPHY:    1     2     3     6
Counties

 Selected          State                   FIPS   press Ctrl-W-----
 ----------------------------------------------
            ALABAMA                | AL |   01
            ALASKA (no date)       | AK |   02
            ARIZONA                | AZ |   04
            ARKANSAS               | AR |   05
        COUNTIES:   (1) - All   (2) - Individual   (R)etur
        All Counties for Selected State
            DISTRICT OF COLUMBIA   | DC |   11
            FLORIDA                | FL |   12
 ----------------------------------------------

 ----- Select/Change/Deselect  -  F1--------------------------------
 Page Up   -  F3     OR     Up/Down   |   SAVE & close window   -  F2
 Page Down  -  F4          arrow keys |   ABORT & close window  -  Esc

 YOU ARE HERE  ->  Main: Search: Criteria: Geography: Counties
```

4



```
GEOGRAPHY:    1     2     3     6
Counties                     CONNECTICUT
                             Selected       County            FIPS
 Selected          State     -----------------------------------------
                                 .   |   FAIRFIELD       |   001
           ALABAMA               .   |   HARTFORD        |   003
           ALASKA (no date)      .   |   LITCHFIELD      |   005
           ARIZONA                   |   MIDDLESEX       |   007
           ARKANSAS                  |   NEW HAVEN       |   009
           CALIFORNIA                |   NEW LONDON      |   011
           COLORADO              .   |   TOLLAND         |   013
           DELAWARE                  |   WINDHAM         |   015
           DISTRICT OF COLUMBI
           FLORIDA

 ----- Select/Change/Deselect  -  F1--------------------------------
 Page Up   -  F3     OR     Up/Down   |   SAVE & close window   -  F2
 Page Down  -  F4          arrow keys |   ABORT & close window  -  Esc

 YOU ARE HERE  ->  Main: Search: Criteria: Geography: Counties
```

Aqueduct Files

USDA Crop Estimates - County file.
Source: US Department of Agriculture, Statistical Reporting Board.
67 commodities reported to the county level, 20 years of data.

National Resources Inventory.
Source: US Department of Agriculture, Soil Conservation Service
1987 inventory of non-Federal lands throughout the United States. 800,000 samples, over 100 fields describing land use, erosion, soil and water types.

Toxic Release Inventory (TRI).
Source: U.S. Environmental Protection Agency
Reports from companies releasing significant amounts of @300 toxic chemicals released into the environment as a result of manufacturing (either as byproduct of process, or actually generated.)

Futures/Options trading – Historical.
Source: Chicago Board of Trade
Trading records (usually monthly) for the commodity markets of the Chicago Board of Trade.

Health and Nutrition Examination Survey (HANES).
Source: National Center for Health Statistics
National level sample (28,000+) surveyed 1976-80 to "measure and monitor the nutritional status and health of the U.S. population ages 6 months through 74 years."

Aqueduct

Project Timeline

Marijo Wilson, Mann Library
Cornell University, Ithaca, New York 14853
STS Forum, ALA 1990

Applied for grant - May 1988

Jan. 1989 April 1989 July 1989 Oct. 1989 Jan. 1990 April 1990 July 1990 Oct. 1990

Grant began – January 1989

First programmer – Jan.-May 1989

First statistician – Jan.-Feb. 1989

Minicomputer selected and installed – Feb.-Aug. 1989

Files selected, sponsors recruited – June - Dec. 1989

Second Programmer –Sept. 1989+

Relational database software selected and installed – Sept. 1989-Feb. 1990

Second statistician – Dec. 1989+

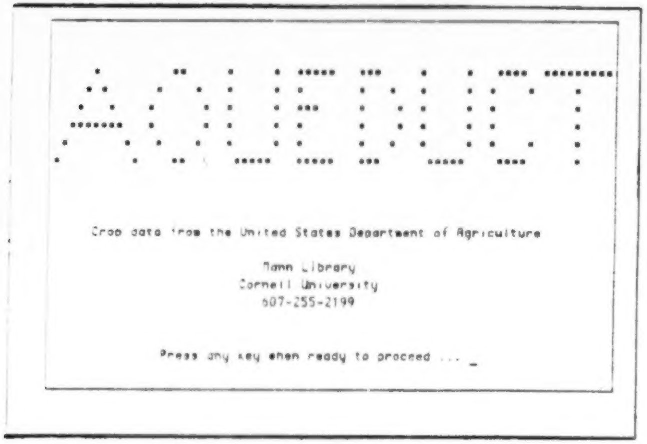
Phase I: basic retrieval system, one file, all fields output. – Feb.-May 1990

Phase II: Complex file, more flexibility, selected fields for output – June -Aug. 1990

Phase III: Two files through one interface – Sept.-Oct. 1990

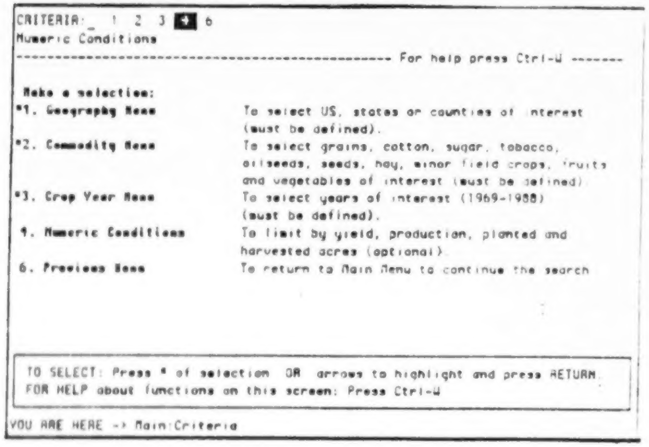
Phase IV: Add more files to system, campus access – Nov.-Dec. 1990

AQUEDUCT - Sample screens - USDA Crop Estimates File



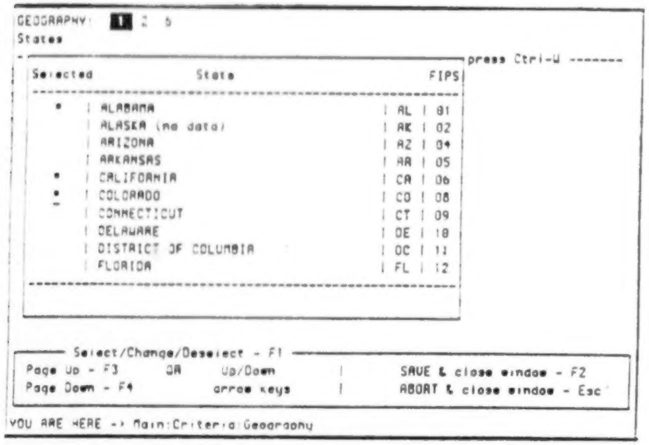
```
Crop data from the United States Department of Agriculture

Mann Library
Cornell University
607-255-2199

Press any key when ready to proceed ... _
```

1. Introductory screen.



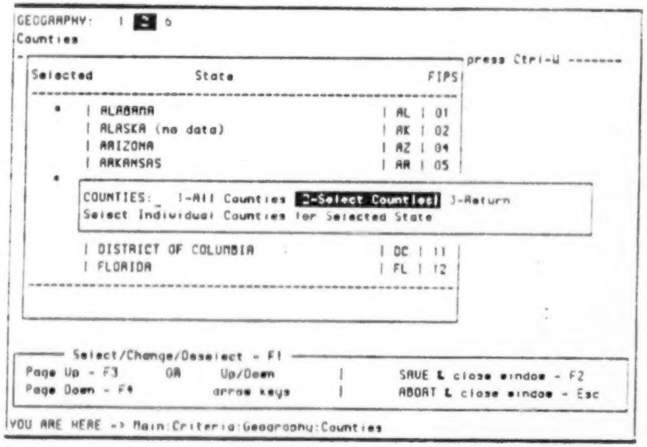
```
CRITERIA: _  1  2  3  4  6
Numeric Conditions
----------------------------------------------- For help press Ctrl-W -------

 Make a selection:
*1. Geography Menu            To select US, states or counties of interest
                              (must be defined).
*2. Commodity Menu            To select grains, cotton, sugar, tobacco,
                              oilseeds, seeds, hay, minor field crops, fruits
                              and vegetables of interest (must be defined).
*3. Crop Year Menu            To select years of interest (1969-1988)
                              (must be defined).
 4. Numeric Conditions        To limit by yield, production, planted and
                              harvested acres (optional).
 6. Previous Menu             To return to Main Menu to continue the search

 TO SELECT: Press # of selection  OR  arrows to highlight and press RETURN.
 FOR HELP about functions on this screen: Press Ctrl-W
YOU ARE HERE -> Main:Criteria
```

2. Menu for search criteria selection.



```
GEOGRAPHY:   1  2  6
States
-                                                     press Ctrl-W -------
 Selected              State                         FIPS
 ---------------------------------------------------------
    *    | ALABAMA                            | AL | 01 |
         | ALASKA (no data)                   | AK | 02 |
         | ARIZONA                            | AZ | 04 |
         | ARKANSAS                           | AR | 05 |
    *    | CALIFORNIA                         | CA | 06 |
    *    | COLORADO                           | CO | 08 |
    _    | CONNECTICUT                        | CT | 09 |
         | DELAWARE                           | DE | 10 |
         | DISTRICT OF COLUMBIA               | DC | 11 |
         | FLORIDA                            | FL | 12 |
 ---------------------------------------------------------

 ------ Select/Change/Deselect - F1 ----------------------------------
 Page Up - F3       OR     Up/Down      |     SAVE & close window - F2
 Page Down - F4            arrow keys   |     ABORT & close window - Esc
YOU ARE HERE -> Main:Criteria:Geography
```

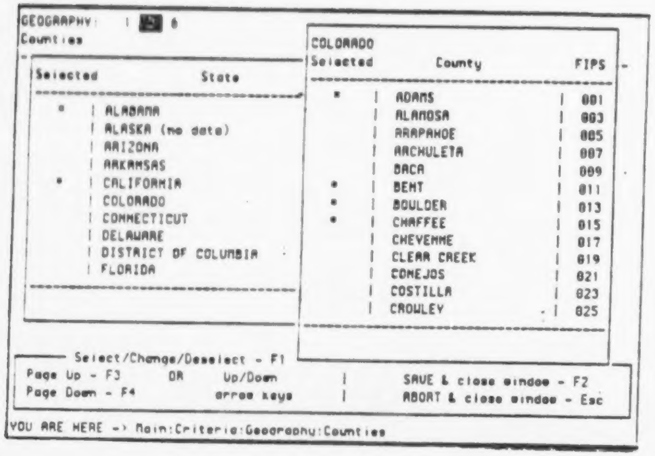
3. Geography search criteria selection
(asterisks indicate selected states)



```
GEOGRAPHY:   1  2  6
Counties
-                                                     press Ctrl-W -------
 Selected              State                         FIPS
 ---------------------------------------------------------
    *    | ALABAMA                            | AL | 01 |
         | ALASKA (no data)                   | AK | 02 |
         | ARIZONA                            | AZ | 04 |
         | ARKANSAS                           | AR | 05 |
    *
      COUNTIES: _  1-All Counties  2-Select Count(ies)  3-Return
      Select Individual Counties for Selected State
         | DISTRICT OF COLUMBIA               | DC | 11 |
         | FLORIDA                            | FL | 12 |
 ---------------------------------------------------------

 ------ Select/Change/Deselect - F1 ----------------------------------
 Page Up - F3       OR     Up/Down      |     SAVE & close window - F2
 Page Down - F4            arrow keys   |     ABORT & close window - Esc
YOU ARE HERE -> Main:Criteria:Geography:Counties
```

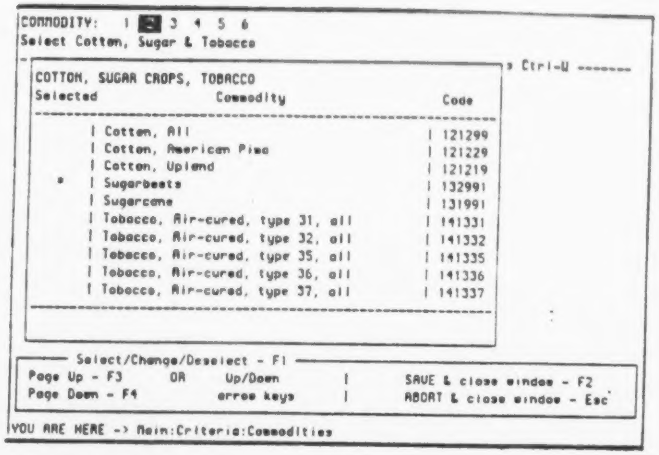
4. Geography selection at county level.

AQUEDUCT - Sample screens - USDA Crop Estimates File (cont'd)



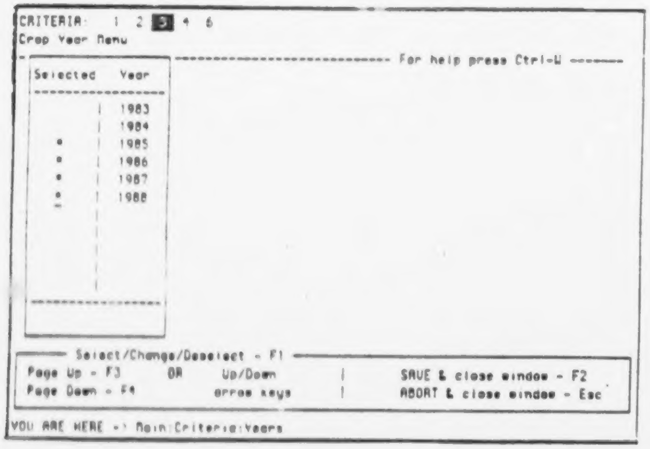
```
GEOGRAPHY:    1 5 6
Counties
                                          COLORADO
 Selected            State                Selected        County          FIPS
 ---------------------------------------  ----------------------------------------
    *    | ALABAMA                           *      | ADAMS               | 001
         | ALASKA (no data)                         | ALAMOSA             | 003
         | ARIZONA                                  | ARAPAHOE            | 005
         | ARKANSAS                                 | ARCHULETA           | 007
    *    | CALIFORNIA                               | BACA                | 009
         | COLORADO                          *      | BENT                | 011
         | CONNECTICUT                       *      | BOULDER             | 013
         | DELAWARE                          *      | CHAFFEE             | 015
         | DISTRICT OF COLUMBIA                     | CHEYENNE            | 017
         | FLORIDA                                  | CLEAR CREEK         | 019
                                                    | CONEJOS             | 021
                                                    | COSTILLA            | 023
                                                    | CROWLEY             | 025

 ----- Select/Change/Deselect - F1
 Page Up - F3      OR     Up/Down        |       SAVE & close window - F2
 Page Down - F4           arrow keys     |       ABORT & close window - Esc
YOU ARE HERE -> Main:Criteria:Geography:Counties
```

5. Geography selection at county level continued.



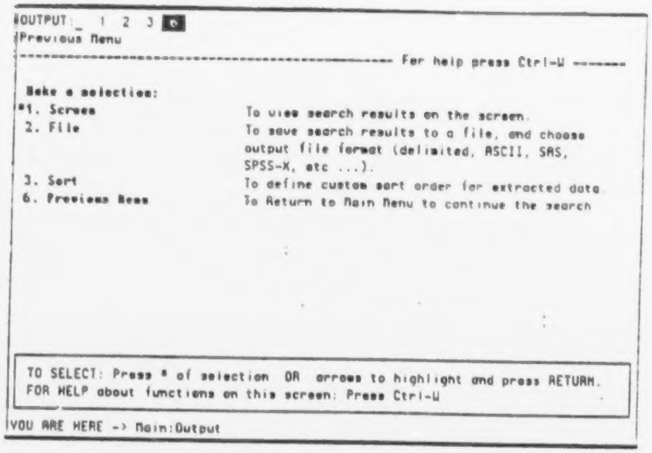
```
COMMODITY:    1 2 3 4 5 6
Select Cotton, Sugar & Tobacco
                                                              > Ctrl-W -------
 COTTON, SUGAR CROPS, TOBACCO
 Selected                   Commodity                      Code
 ---------------------------------------------------------------
         | Cotton, All                                   | 121299
         | Cotton, American Pima                         | 121229
         | Cotton, Upland                                | 121219
    *    | Sugarbeets                                    | 132991
         | Sugarcane                                     | 131991
         | Tobacco, Air-cured, type 31, all              | 141331
         | Tobacco, Air-cured, type 32, all              | 141332
         | Tobacco, Air-cured, type 35, all              | 141335
         | Tobacco, Air-cured, type 36, all              | 141336
         | Tobacco, Air-cured, type 37, all              | 141337
 ---------------------------------------------------------------

 ----- Select/Change/Deselect - F1 -----
 Page Up - F3      OR     Up/Down        |       SAVE & close window - F2
 Page Down - F4           arrow keys     |       ABORT & close window - Esc
YOU ARE HERE -> Main:Criteria:Commodities
```

6. Commodity search criteria selection menu.



```
CRITERIA:   1 2 3 4 6
Crop Year Menu
- Selected   Year   ---------------------------- For help press Ctrl-W -------
 ----------------
           | 1983
           | 1984
    *      | 1985
    *      | 1986
    *      | 1987
    *      | 1988
    -      |
           |
 ----------------

 ----- Select/Change/Deselect - F1 -----
 Page Up - F3      OR     Up/Down        |       SAVE & close window - F2
 Page Down - F4           arrow keys     |       ABORT & close window - Esc
YOU ARE HERE -> Main:Criteria:Years
```

7. Crop year search criteria selection menu.

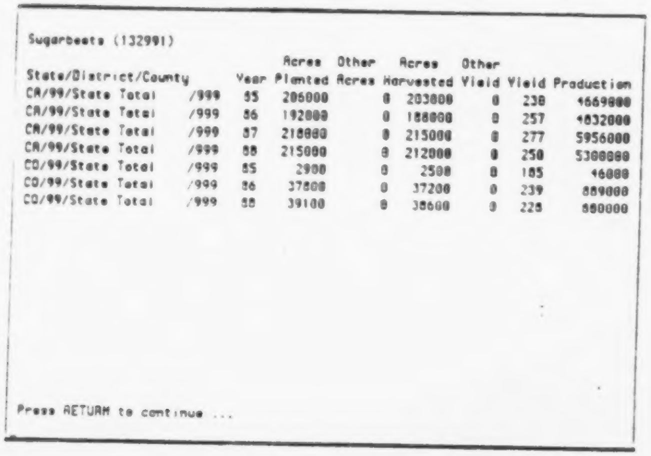


```
OUTPUT:_   1 2 3 6
Previous Menu
------------------------------------------------ For help press Ctrl-W -------

 Make a selection:
*1. Screen                  To view search results on the screen.
 2. File                    To save search results to a file, and choose
                            output file format (delimited, ASCII, SAS,
                            SPSS-X, etc ...).
 3. Sort                    To define custom sort order for extracted data.
 6. Previous Menu           To Return to Main Menu to continue the search

 TO SELECT: Press # of selection  OR  arrows to highlight and press RETURN.
 FOR HELP about functions on this screen: Press Ctrl-W
YOU ARE HERE -> Main:Output
```

8. Output format selection menu.

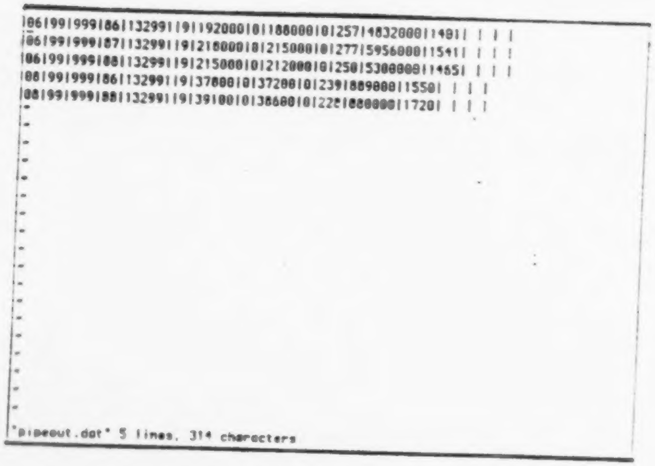
BEST COPY AVAILABLE

AQUEDUCT - Sample screens - USDA Crop Estimates File (cont'd)



```
Sugarbeets (132991)
                                     Acres  Other    Acres  Other
State/District/County          Year Planted Acres Harvested Yield Yield Production
CA/99/State Total      /999     85   206000     0   203000     0   230    4669000
CA/99/State Total      /999     86   192000     0   188000     0   257    4832000
CA/99/State Total      /999     87   218000     0   215000     0   277    5956000
CA/99/State Total      /999     88   215000     0   212000     0   250    5300000
CO/99/State Total      /999     85     2900     0     2500     0   185      46000
CO/99/State Total      /999     86    37800     0    37200     0   239     889000
CO/99/State Total      /999     88    39100     0    38600     0   228     880000

Press RETURN to continue ...
```

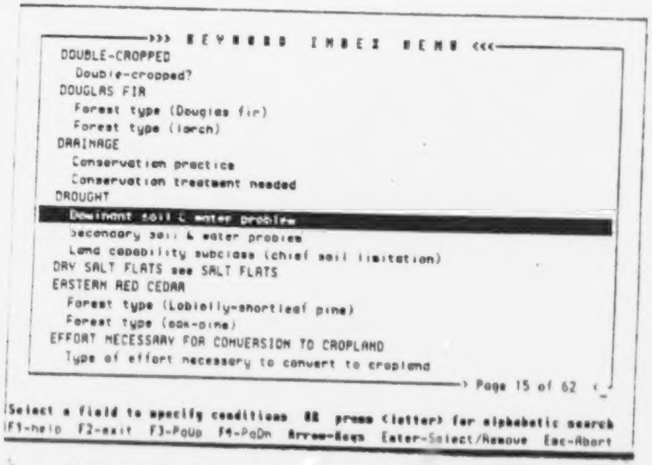
9. Output formated for screen display.



```
|06|99|999|86||13299||9||192000|0||188000|0|257|4832000||148| | | |
|06|99|999|87||13299||9|218000|0|215000|0|277|5956000||154| | | |
|06|99|999|88||13299||9|215000|0|212000|0|250|5300000||465| | | |
|08|99|999|86||13299||9|37800|0|37200|0|239|889000||1550| | | |
|08|99|999|88||13299||9|39100|0|38600|0|228|880000||1720| | | |
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
"pipeout.dat" 5 lines, 314 characters
```

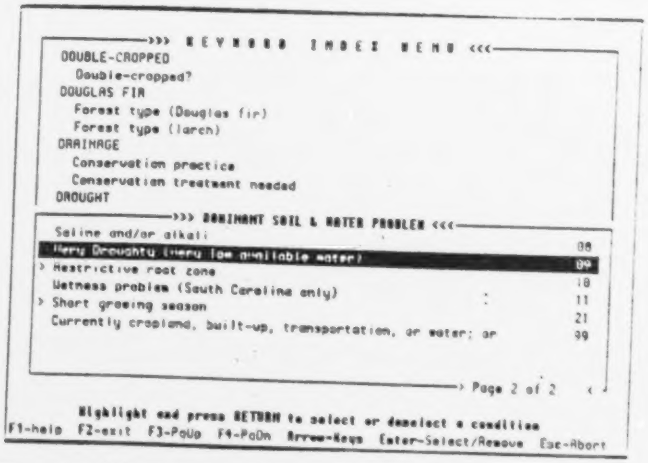
10. Formatted output sent to file for manipulation.

AQUEDUCT - Sample screens - National Resources Inventory



```
 ------>>> KEYWORD INDEX MENU <<<------
  DOUBLE-CROPPED
    Double-cropped?
  DOUGLAS FIR
    Forest type (Douglas fir)
    Forest type (larch)
  DRAINAGE
    Conservation practice
    Conservation treatment needed
  DROUGHT
    Dominant soil & water problem
    Secondary soil & water problem
    Land capability subclass (chief soil limitation)
  DRY SALT FLATS see SALT FLATS
  EASTERN RED CEDAR
    Forest type (Loblolly-shortleaf pine)
    Forest type (oak-pine)
  EFFORT NECESSARY FOR CONVERSION TO CROPLAND
    Type of effort necessary to convert to cropland
 -------------------------------------------> Page 15 of 62 <
Select a field to specify conditions  OR  press <letter> for alphabetic search
F1-help  F2-exit  F3-PgUp  F4-PgDn  Arrow-Keys  Enter-Select/Remove  Esc-Abort
```

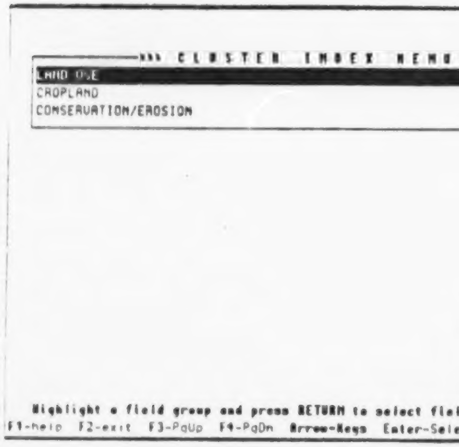
1. Keyword selection menu (reverse video indicates selection of field name under keyword)



```
 ------>>> KEYWORD INDEX MENU <<<------
  DOUBLE-CROPPED
    Double-cropped?
  DOUGLAS FIR
    Forest type (Douglas fir)
    Forest type (larch)
  DRAINAGE
    Conservation practice
    Conservation treatment needed
  DROUGHT
 -------->>> DOMINANT SOIL & WATER PROBLEM <<<------
   Saline and/or alkali                                            08
   Very Droughty (very low available water)                        09
 > Restrictive root zone                                           10
   Wetness problem (South Carolina only)                           11
 > Short growing season                                            21
   Currently cropland, built-up, transportation, or water; or      99
 -------------------------------------------> Page 2 of 2 <
        Highlight and press RETURN to select or deselect a condition
F1-help  F2-exit  F3-PgUp  F4-PgDn  Arrow-Keys  Enter-Select/Remove  Esc-Abort
```

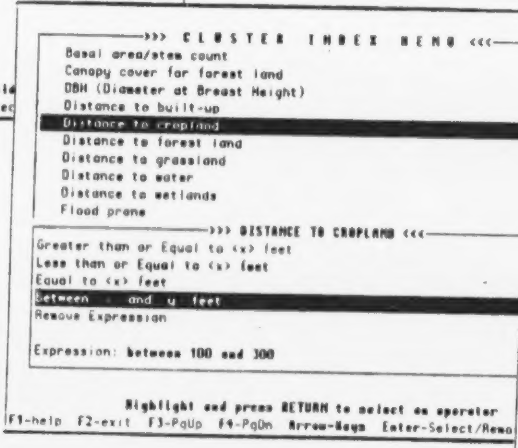
2. Further selection of variables associated with selected field name.

AQUEDUCT - Sample screens - National Resources Inventory (cont'd)



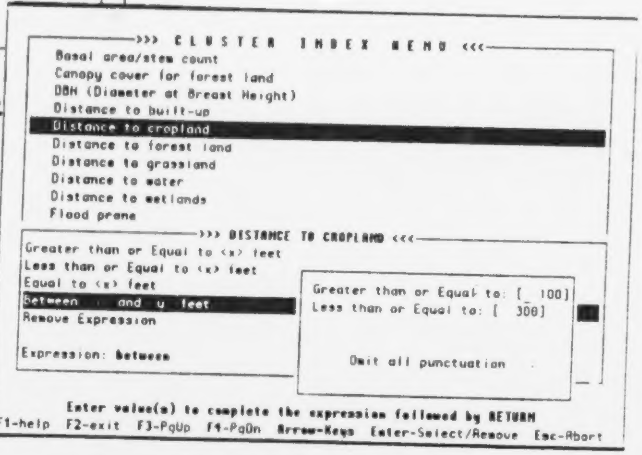
```
——>>> CLUSTER INDEX MENU <<<——
LAND USE
CROPLAND
CONSERVATION/EROSION

Highlight a field group and press RETURN to select field
F1-help  F2-exit  F3-PgUp  F4-PgDn  Arrow-Keys  Enter-Selec
```

3. Cluster Index menu (field names are accessible via 3 broad groupings)



```
——>>> CLUSTER INDEX MENU <<<——
Basal area/stem count
Canopy cover for forest land
DBH (Diameter at Breast Height)
Distance to built-up
Distance to cropland
Distance to forest land
Distance to grassland
Distance to water
Distance to wetlands
Flood prone
——>>> DISTANCE TO CROPLAND <<<——
Greater than or Equal to <x> feet
Less than or Equal to <x> feet
Equal to <x> feet
Between x and y feet
Remove Expression

Expression: between 100 and 300

Highlight and press RETURN to select an operator
F1-help  F2-exit  F3-PgUp  F4-PgDn  Arrow-Keys  Enter-Select/Remo
```

4. Selection of field names in Land Use Cluster.



```
——>>> CLUSTER INDEX MENU <<<——
Basal area/stem count
Canopy cover for forest land
DBH (Diameter at Breast Height)
Distance to built-up
Distance to cropland
Distance to forest land
Distance to grassland
Distance to water
Distance to wetlands
Flood prone
——>>> DISTANCE TO CROPLAND <<<——
Greater than or Equal to <x> feet
Less than or Equal to <x> feet      Greater than or Equal to: [  100]
Equal to <x> feet                   Less than or Equal to: [  300]
Between x and y feet
Remove Expression

Expression: between                   Omit all punctuation

Enter value(s) to complete the expression followed by RETURN
F1-help  F2-exit  F3-PgUp  F4-PgDn  Arrow-Keys  Enter-Select/Remove  Esc-Abort
```

5. Menu for definition of numeric conditions for a selected field.

BEST COPY AVAILABLE